DOCUMENT RESUME

ED 432 371 PS 027 705

AUTHOR	Friedman, Mark
TITLE	A Guide to Developing and Using Performance Measures in Results-Based Budgeting.
INSTITUTION	Finance Project, Washington, DC.
PUB DATE	1997-05-00
NOTE	52p.
AVAILABLE FROM	Finance Project, 1000 Vermont Avenue, NW, Suite 600, Washington, DC 20005; Tel: 202-628-4200; Fax: 202-628-4205 ($7.50).
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Accountability; *Budgeting; Educational Finance; *Evaluation Criteria; *Financial Support; Performance Factors; Planning; Public Agencies; *Public Education; Strategic Planning
IDENTIFIERS	*Fiscal Accountability; *Indicators

ABSTRACT

Intended as a guide for taking steps to hold children's project initiatives accountable for their performance by meeting reasonable targets while maintaining a focus on overall desired results, this report was developed to create performance accountability within a results framework. Characteristics of an effective performance measurement system are presented, along with examples of state and local performance measurement systems. Also discussed is a four-quadrant approach for developing performance measures that are useful, quantifiable, and that fit into a more broad performance measurement framework. The report stresses the crucial distinction between two types of accountability, accountability for results and that for performance: results accountability concerns conditions of well-being for children, families, and communities that cut across agencies and programs, whereas performance accountability is that part of results accountability concerned with how well agencies and programs perform. Taken together, these two levels of accountability cover the whole range of questions, from the broadest level view of community accountability for child and family well-being to the smallest increment of performance by a particular program. Along with an introduction and conclusion, the report has six sections: (1) starting points; (2) an approach to performance measurement; (3) characteristics of an effective performance measurement system; (4) using performance measures to improve performance; (5) examples of state and local performance measurement systems; and (6) lessons and issues. Contains 42 references. (LBT)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 432 371

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A GUIDE TO DEVELOPING AND USING PERFORMANCE MEASURES IN RESULTS-BASED BUDGETING

■

PS 027205

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Cheryl D. Hayes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

THE FINANCE PROJECT

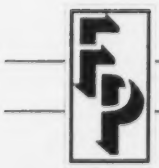
BEST COPY AVAILABLE

A GUIDE TO DEVELOPING AND USING PERFORMANCE MEASURES IN RESULTS-BASED BUDGETING

May 1997

By Mark Friedman

Prepared For
THE FINANCE PROJECT
1000 Vermont Avenue, NW
Washington, DC 20005
202/628-4200
Fax: 202/628-4205



THE FINANCE PROJECT

ABOUT THE AUTHOR

Mark Friedman served for 19 years in the Maryland Department of Human Resources, including six years as the Department's chief financial officer. After four years with the Center for the Study of Social Policy, he established and now directs the Fiscal Policy Studies Institute in Baltimore, Maryland. Mark is a member of The Finance Project's Working Group on Results-based Planning, Budgeting, Management, and Accountability Systems.

PREFACE

Public financing for education and an array of other children's services has become a topic of significant interest and political concern. Growing skepticism among a critical mass of American voters and taxpayers has fueled doubts about the ability of government to solve problems and provide basic supports and services that enhance the quality of life in their communities. Many believe government is too big, that it's too expensive, and that it doesn't work very well.

Despite steadily increasing public expenditures for health, education, welfare, human services, and public safety over the past two decades, seemingly intractable problems persist. Nearly a quarter of the children in the U.S. are poor and live in families and communities that are unable to meet their basic needs. Schools have become increasingly expensive, but student achievement has not matched the rising costs and dropout rates remain unacceptably high. Health care costs continue to go up, yet many Americans can't get the services they need, and with each passing year their health care dollars buy less. Criminal justice demands a dramatically increasing share of public dollars—for police officers, judges, and jails—but neighborhood streets don't seem any safer.

Voters have spoken clearly. They want more for their money—more and better services, yes, but also balanced budgets and cuts in income and property taxes. After more than a decade of chronic deficits, they want government at all levels to operate more effectively and efficiently. They don't want to dismantle government, but rather they want government to meet vital public needs and make a more visible difference in their lives.

Elected officials and other policy makers have responded to public concern and dissatisfaction by focusing more explicitly on the results of the programs and initiatives that they develop and fund. Reformers have sought to redefine the missions of public programs and agencies, to modify how services are delivered, to measure how well government programs and agencies are performing, and to feed information about performance back into planning, budgeting, management, and accountability systems. While the federal government's National Performance Review and its initiatives to "reinvent government" may be the most prominent examples of this focus on results, there are countless other efforts at the state and local levels that span the divisions of ideology, political party, and the executive and legislative branches of government.

Focusing on results is particularly important for programs and policies serving children and their families. The future well-being of the nation is obviously tied to children's healthy development. Yet policy makers and citizens alike may be inclined to reduce their commitment to critical supports and services without strong evidence that these investments yield results that society cares about, such as healthy children, children succeeding in school, strong families, and safe homes and neighborhoods.

Unfortunately, many of the efforts to implement a results framework—for public programs generally, as well as those targeted to children and their families—have been marred by confusion about terms and basic definitions, insufficient political understanding and support, the difficulty of identifying appropriate results and performance measures, and the challenges of overhauling existing planning, budgeting, and management systems. Policy makers trying to implement results-based systems have enthusiastically set out in many different directions, but often without a particular destination or a map to help them get there.

The Finance Project, established by a consortium of national foundations, conducts an ambitious agenda of policy research and development activities to improve the effectiveness, efficiency, and equity of public financing for education and other children's services. Among these efforts, is assisting with the important work of achieving and measuring important outcomes for children, their families, and the communities in which they live. To guide its work

in this area, The Finance Project created a Working Group on Results-Based Planning, Budgeting, Management, and Accountability Systems.

Under the direction of the working group, a *Strategy Map for Results-Based Budgeting* was designed as a road map for those desiring to incorporate results in their planning and budgeting systems. The *Strategy Map* defines results, indicators, and performance measures and offers a framework for choosing them. It describes the products and competencies required for designing and putting into place a results-oriented budgeting system and discusses lessons from existing initiatives to define, measure, and achieve results. It suggests how to build political and community support, how to reallocate resources and tie them to results, how to integrate results-based budgeting into an existing budgeting process, and how to avoid common pitfalls. It serves as a framework for a series of papers and tool kits for creating results-based planning and budgeting systems that are under development by The Finance Project: a guide to results and indicators, a guide to performance measures, a tool kit on children's budgets, and a paper presenting a cost-of-failure/cost-of-bad-results prototype and analysis.

This paper, *A Guide to Developing and Using Performance Measures in Results-based Budgeting*, is one of the tools that the *Strategy Map* spawned. It is intended as a user-friendly guide for taking steps to hold initiatives accountable for their performance by meeting reasonable targets while maintaining a focus on overall desired results. It is a guide to creating performance accountability within a results framework. It presents characteristics of an effective performance measurement system, examples of state and local performance measurement systems, and a "four-quadrant approach" for developing performance measures that are useful, quantifiable, and that fit into a more broad performance measurement framework.

The paper was prepared by Mark Friedman of the Fiscal Policy Studies Institute. He and I would like to recognize Anna Danegger and Jason Juffras, who as Finance Project staff members assisted in research, conceptualization, and production of this paper. And we would like to thank John Barton, Trine Bech, Peggy O'Sullivan Kachel, John Dorman, Ginny McKay, Jolie Bain Pillsbury, Joan Reeves, Verne Skagerberg, and Marv Weidner for their significant assistance. The information that they provided and their helpful and constructive comments are reflected in the paper that follows.

Cheryl D. Hayes
Executive Director

TABLE OF CONTENTS

Preface		i
I.	Introduction	1
II.	Starting Points	2
III.	An Approach to Performance Measurement	4
IV.	Characteristics of an Effective Performance Measurement System	16
V.	Using Performance Measures to Improve Performance	19
VI.	Examples of State and Local Performance Measurement Systems	25
VII.	Lessons and Issues	29
VIII.	Conclusion	34

"Cheshire Cat," Alice began, "Would you tell me, please, which way I ought to go from here?" "That depends a good deal on where you want to get to," said the Cat.

- Lewis Carroll

Hours after the last familiar sign, the driver kept up a steady pace. "We're lost, aren't we?" said the passenger. "Yes," said the driver, "But we're making good time, don't you think?"

- Anon.

"Thank God we don't get the government we pay for."

- Will Rogers

I. INTRODUCTION

Will Rogers' cynicism about the performance of government still captures a common, if not always constructive, part of public life at the end of the 20th century. And as contract relationships blur the boundaries between the public and private sectors, confidence in private-sector programs has eroded as well, sometimes as guilt by association. The toll is arguably highest among programs that provide health, education, and social services for families and children. If the public is right, if the performance of these programs is not what it should be, then how can we do better? And before we answer that question, how do we know that we are doing badly? **How do we know what "better" is?** This paper is about answering these common-sense questions. It addresses the art of knowing whether our programs and agencies are succeeding or failing, and how to use performance accountability to improve performance.

The title of this paper contains a crucial distinction between two types of accountability: accountability for **results** and accountability for **performance**. Results accountability deals with conditions of well-being for children, families, and communities that cut across agencies and programs.[1] Performance accountability is that part of results accountability concerned with how well agencies and programs perform. Taken together, these two levels of accountability cover the whole range of questions from the broadest-level view of community accountability for child and family well-being to the smallest increment of performance by a particular program (and even a particular individual).[2]

This paper is part of a series of papers published by The Finance Project on the subject of results accountability. *A Strategy Map for Results-based Budgeting* addresses what a results-based budgeting system might look like and how to begin to put it in place. This paper addresses the challenge, **embedded** in the first, of how to hold programs accountable for the

[1] While this paper uses examples primarily from the realm of child and family services, the concepts can be applied to public and private services of all types.

[2] See *From Outcomes to Budgets*, July 1995, from the Center for the Study of Social Policy, Washington, D.C., and *A Strategy Map for Results Based Budgeting*, September 1996, from The Finance Project, Washington, D.C., for a more complete discussion of results-based decision making and budgeting.

best possible performance, while ensuring that their performance is aligned with, and supports, overall efforts to improve results—in other words, how to create performance accountability within a results framework.

II. STARTING POINTS

We start with a few conventions[3] that will help us work on performance measurement in a clear and disciplined way.

A. First, Words about Language (again)

There is an astounding lack of discipline in the use of language in the current work on child and family well-being. It is quite common to find people using the same terms in different, sometimes contradictory, ways, and then wondering why they are not making progress. Processes without a common language tend to be frustrating and unproductive.

The following definitions provide the conceptual starting point for our discussion of results and performance accountability.

> Result (or Outcome)[4]: A "result" is a **bottom line** condition of well-being for children, families, or communities. Results are matters of common sense, above and beyond the jargon of bureaucracy. They are about the fundamental interests of citizens and the fundamental purposes of government. Results are, by definition, not "owned" by any single agency or system. They cross over agency and program lines, and public and private sectors. Examples of results include: children born healthy, children ready for school, children succeeding in school, young people avoiding trouble, stable and self-sufficient families, and safe and supportive communities.
>
> Indicator (or Benchmark)[5]: An "indicator" is a measure, for which we have data, that helps quantify the achievement of a desired result. Indicators help answer the question: "How would we know a result if we achieved it?" Examples of indicators include: rates of preventable disease among children; reading and math achievement scores; high school graduation rates; rates of teen pregnancy and drug use; and crime rates.
>
> Performance measure: A "performance measure" is a measure of how well public or private agencies and programs are working. Typical performance measures address matters of timeliness, cost-effectiveness, and compliance with standards. Examples of performance measures include: percentage of child abuse investigations initiated within 24 hours of a report; amount of child support collected for each dollar expended on child support enforcement; and police or fire response time.

---

[3] Other conventions available in earlier papers: Shriners, Geneva, Blackwood.

[4] In some parts of the country, the term "outcome" has taken on a political meaning very different from the way in which we use the term here. Our use of "outcome" to mean a condition of well-being for children, families, or communities stands in contrast to its usage in the outcome-based education debate, where the term describes new approaches to measuring a student's knowledge and skills. For this reason, we will give preference to the term "result" in the sections that follow.

[5] Note the difference in the way in which the term "benchmark" is used in public- and private-sector applications. The public sector often uses the term "benchmark" to mean an indicator or performance measure. The private sector uses the term to mean a particular level of (desired and achievable) performance. See the discussion that follows in Section V-E.

Performance measures are absolutely essential for running programs well. But they are very different from results and indicators. They have to do with our service **response** to social problems, not the **conditions** that we are trying to improve. It is possible, even common, for individual programs to be successful, while overall conditions get worse.

The key distinction in this set of definitions is between ends and means. Results and indicators have to do with ends. Performance measures and the programs they describe have to do with means. The end we seek is not "better service"[6] but better results. These distinctions will help us to describe budgeting processes built on clear thinking about what we wish to achieve and the strategies that we choose to get there.

B. The Change-Agent vs. Industrial Model of Services

Much of the tradition of performance measurement comes from the private sector and, in particular, the industrial part of the private sector. Work measurement—dating back to the time and motion studies of the late 19th and early 20th centuries—looked at how to improve production. Industrial processes turn raw materials into finished products. The raw materials are the inputs; the finished products are the outputs.

This model does not translate very well into public or private sector enterprises that provide **services**.[7] It does not make much sense to think of clients, workers, and office equipment as inputs to the service sausage machine, churning out served, cured, or fixed clients. Instead, we need to begin thinking about services in terms of the **change-agent model**. The agency or program provides services (inputs) that act upon the environment to produce demonstrable changes in the well-being of clients, families, or communities (outputs).

One common situation illustrates the problems that arise when industrial-model thinking is applied to services. It is the belief that the number of clients served is an output. ("We have assembled all these workers in all this office space, and we are in the business of processing unserved clients into served clients.") This misapplication of industrial-performance concepts to services captures much of what is wrong with the way we measure human-service performance today. "Number of clients served" is **not an output**. It is an input, an action that should lead to a change in client or social conditions—the real output we are looking for. ["We served 100 clients (input) and 50 of them got jobs (output) and 40 of them still had jobs a year later (even more important output)."] This is a whole different frame of mind and a whole different approach to performance measurement.

A closely related industrial-model problem involves treating dollars spent as inputs, and clients served as outputs. In this distorted view, dollars are raw materials, and whatever the program happens to do with those dollars are outputs. It is easy to see why this oversimplification fails to meet the public's need for accountability. In this construct, the mere fact that the government spent all the money it received is a type of performance measurement. This is surely a form of intellectual, and perhaps literal, bankruptcy. In this

---

[6] Or even "integrated service." Service integration is a means, not an end in itself.

[7] It is important to note that performance work in the private sector, including the "industrial" sector, has gone beyond the simple model noted here. In particular, the growing corporate service sector has many companies that have successfully addressed the challenges discussed in this paper. The intent here is not to set up industrial models and measurements as straw men, but to suggest that some public and private agencies are still stuck with performance models that do not work well for service organizations.

perverse scheme, almost all the agency's data are purportedly about outputs. This gives the agency the appearance of being output-oriented and very progressive. It just doesn't happen to mean anything.

Much of the confusion about performance measurement derives from the attempt to impose industrial-model concepts on change-agent services. The best model would be one that could span industrial and change-agent applications. Some government services still involve industrial-type production (although these are often the best candidates for privatization and a diminishing breed.) In other cases, the service itself (or components of the service) has product-like characteristics, and industrial model concepts apply well. But most government and private-sector human services fall into the change-agent category. We will concentrate the following discussion on services that fit the change-agent model, but the approach described in Section III can be used for either industrial or change-agent applications.

C. Point of View

Finally, as you may have guessed by now, this is a paper with a point of view. It is not a neutral summary of work in the field. It does not hold all performance measures or measurement systems to be created equal. Rather, it proposes a way to approach performance measurement in what we hope will be a clear, common-sense, and, most of all, useful way. What may be seen as implied criticism of other approaches is not intended to diminish the value of this other work, or to set up the approach offered here as inherently superior. The business of public accountability is extraordinarily difficult and often thankless work. The states, counties, and communities referenced here deserve great credit for their efforts. Only by trying things and learning from each other will we have a chance to make measurable progress on performance accountability.

III. AN APPROACH TO PERFORMANCE MEASUREMENT

In this section, we offer an approach to performance measurement. While this is only one of many possible approaches, we think that it is worth consideration for two reasons:

First, it aligns precisely with the results-based decision-making and budgeting framework presented in earlier work. This means that these two pieces together provide an approach to accountability that spans the distance from the highest-level view of the well-being of children and families (across agencies and across communities) to the lowest-level view of how individual programs (and even individuals within organizations) perform.

Second, this approach to performance measurement can be used to assess other performance measurement systems for completeness. We believe that the four-quadrant approach to developing performance measures described below provides a framework that accounts for the way that most, if not all, performance measurement systems fit together.

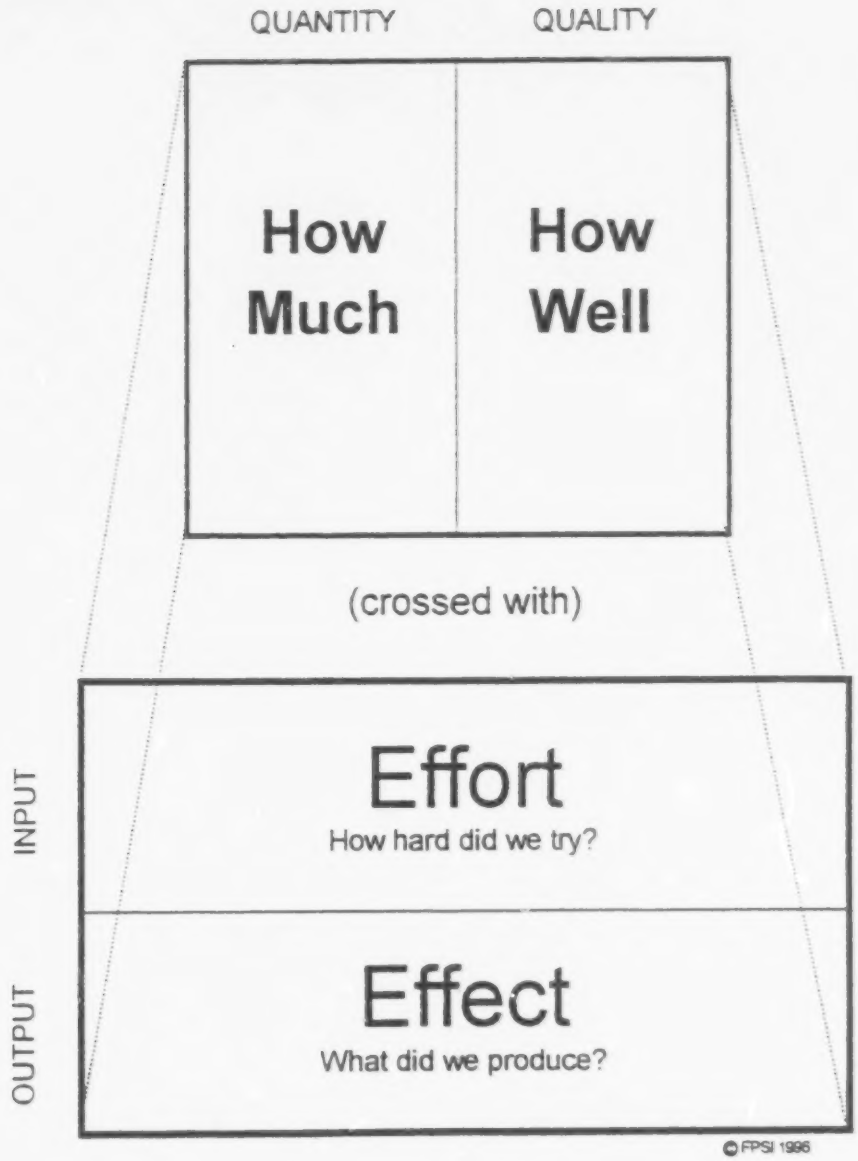
A. The Four-Quadrant Approach to Performance Measurement

A lens through which to view the field of performance measurement

The heart of any performance measurement system is the way in which data are categorized, selected, and used. The various approaches to performance measurement have produced different ways of doing this. In this section, we offer a scheme for categorizing and selecting performance measures. In Section IV we discuss the characteristics of an effective performance measurement system. And in Section V we address the matter of how performance measures can be used.

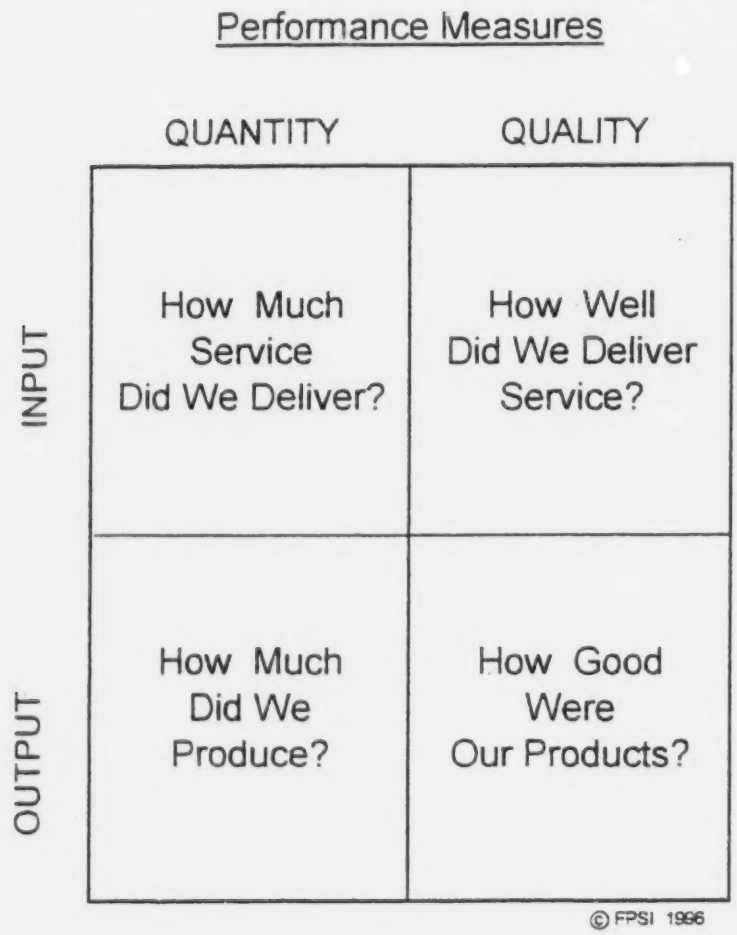
Let's cut this problem down to its bare essentials: how do we choose **data elements** to measure performance? If we can answer this question, much of the rest follows suit.

All work on performance measurement tries to answer two sets of interlocking questions:



We therefore reach the following bold...

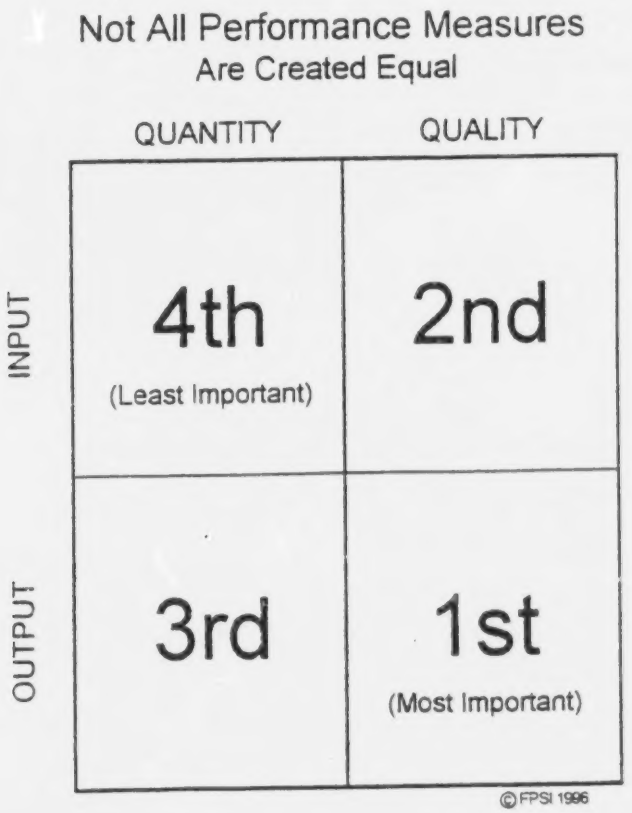
Assertion: All performance measures can be sorted into four categories, represented by the following four-quadrant matrix:



This sorting scheme allows us to pose and answer some common sense questions about performance. These are shown in their most basic form in the chart on the previous page.

- **Upper-left quadrant: How much service did we deliver?** How much effort did we put into service delivery? How hard did we try?
- **Upper-right quadrant: How well did we deliver service?** How well did we treat our customers? Was service courteous, timely, accessible, consistent, etc.?
- **Lower-left quadrant: How much did we produce?** How many clients or customers showed an improvement in well-being? How much do we have to show for our service?
- **Lower-right quadrant: How good were our products?** What percentage of our clients or customers showed improvement? What do we have to show for our service in terms of output quality?

One of the immediate consequences of this sorting scheme is that **not all of these questions are equally important**. We are (or we should be) far more interested in quality than in quantity. And it is not enough to count effort; we must also measure effect.



Many performance measurement documents provide a great deal of information on quantity of input (upper left), but very little on quality and output (the other three quadrants). Performance measures tend to deal exclusively with how many clients were served, how many applications were processed, etc. In some cases, these systems put forward even less appropriate industrial-model quantity measures, such as "how many workers do we have, how much space, how much money, etc.," not how much was produced, and how well.

This matrix allows us to separate the wheat from the chaff in selecting performance measures. Performance measurement should focus on the quality column measures and, in particular, on the quality of output measures. Therefore, we can actually assign an order of importance to the four quadrants as shown above. We need to move from our preoccupation with the upper-left quadrant, toward the upper- and lower-right quadrants.

B. "Get to the Point"[8] Planning

Notice how we have skipped right past mission, vision, values, purpose, goals, and objectives and gone directly to performance measures. Now, this goes against the orthodoxy of the planning and budgeting profession, but it is possible and even desirable to do this. First, it gets people into the work right away. Second, it gets us past the tyranny of planning systems that decree that the work is linear and that program measurements must somehow be **derived** from higher-level statements of purpose. Baloney.

There is no reason to start with agency mission. It can, in fact, be argued that, by working **down** from results and **up** from programs, agency mission statements become a by-product of this work. Mission statements and their attendants, retainers, and attorneys help articulate why the agency exists—how it contributes to improving results—and generally how it goes about doing this. But there is no reason to wait for the perfect articulation of mission before getting about the business of selecting performance measures.

You can go back and do the mission(ary) stuff later if you want. It is probably a good idea for agencies to be able to state in a few phrases what they are about. But it is unnecessarily time-consuming and burdensome to try to develop performance measures from these statements, as if it is a matter of mathematical derivation. Unless you are thinking of creating a brand new agency, most people who face performance measurement challenges have programs that need performance measurement in practical forms right now.

Think about it this way: results accountability tells us whether a program should exist (or not) as part of our larger strategy to improve ("turn the curve") on child and family well-being.[9] Performance measurement picks up at this point, takes as given that the program needs to be there, and moves to the next step of answering whether it is working or not.

"Traditional" planning systems spend an inordinate amount of time on preliminaries before people actually get to talk about how to measure performance. By going straight to the business of selecting performance measures, we ease the frustration and associated cynicism that go with complex planning processes. We also get to the heart of what may be the benefit of performance measurement, namely, a disciplined way to use data in the day-to-day management of programs.[10]

---

[8] For crying out loud.

[9] See the discussion of "turning the curve" in the "Strategy Map" paper, pages 5-7 and 41.

[10] The Treasury Department survey of major corporations found that 90% of all measures in actual use were "developed as part of some unit-initiated improvement effort." (*Performance Measurement: Report on A Survey of Private Sector Performance Measures*, Department of the Treasury, January 1993, page 11.)

Another benefit of this four-part system is its simplicity and (arguable) common sense. Many performance measurement systems suffer from the creation of so many special terms and variations on special terms that it is hard to keep them straight. (**Ten or more types** of performance measures are not uncommon.) Some of this problem derives from the fact that these systems often do not distinguish between results, indicators, and performance measures, and thus create unnecessary complexity trying to keep this straight. Another related problem comes from an attempt to strictly define how many "levels" there are to a performance system. Some performance systems call performance measures **by different names at different levels of the organization**. This does not work well, because there are varying numbers of layers in different organizations. In the four-quadrant approach, we have a single framework that is repeated, in more or less the same way, through as many levels as exist in a given organization.[11]

C. Examples of Program Performance Measures Using the Four-Quadrant Approach

Following are some examples of performance measures using the four-quadrant approach. For purposes of illustration, we phrase each entry in terms of a question, but, in practice, the **entries for each quadrant are data elements that answer the question.**

Education

	Quantity	Quality
Input	☐ How many students did we serve this year?	☐ What was our student/teacher ratio? ☐ What percent of our teachers have advanced degrees? ☐ How "rich" is our extracurricular program?
Output	☐ How many children graduated? ☐ How many children dropped out?	☐ What percent graduated on time? ☐ What percent completed advanced placement courses? ☐ What percent entered work or college after graduation? ☐ What were average earnings for our students 2 and 5 years later?

[11] For those interested in the parallel to fractal geometry, see *Chaos, Making a New Science*, James Gleick, 1987, page 98.

Health

	Quantity	Quality
Input	☐ How many patients have we served? ☐ How many clients are enrolled?	☐ How long is the wait for an appointment? ☐ How accessible are our offices (% of patients within 20 minute trip from home or school)? ☐ How often do we see children at or near their school? ☐ What percent of children receive well-baby or preventive appointments?
Output	For our client population: ☐ How many acute-care visits? ☐ How many hospital days? ☐ How many preventable illnesses? ☐ How many healthy births?	For our client population: ☐ What percent of children are fully immunized ☐ What percent of births are healthy? (low birthweight % or birth complications %) ☐ What percent of children experience preventable illness?

Child Welfare

	Quantity	Quality
Input	☐ How many foster children did we serve? ☐ How many child abuse investigations did we complete?	☐ How often did children change foster care placement? ☐ How many abuse investigations were initiated within 24 hours? ☐ What is the average length of stay in foster care? ☐ What is the average wait for adoption?
Output	☐ How many foster children were reunified with their natural families? ☐ How many foster children were placed in permanent adoptive homes? ☐ How many child abuse cases resulted in children able to stay safely at home?	☐ What percent repeat abuse reports were received on reunification cases? ☐ What percent of adoptive placements were stable one year, two years later? ☐ What percent of foster children graduated on time from high school? What percent entered the workforce?

Welfare "Reformed"

	Quantity	Quality
Input	☐ How many clients/families did we serve? ☐ How many were placed in job training?	☐ What percent of those served had employment support plan needs met? (e.g. child care, transportation etc.) ☐ What percent of those served were long-term recipient cases?
Output	☐ How many clients successfully completed employment training? ☐ How many were employed in non-subsidized employment?	☐ What percent of clients served were employed? ☐ What was the job retention rate at 6 12, and 24 months? ☐ What percent of jobs had health insurance? ☐ What was the cost/benefit ratio of the employment program? (direct costs vs. reduced/avoided welfare payments) ☐ What was the welfare reentry rate?

Mental Health

	Quantity	Quality
Input	☐ How many service appointments (inpatient and outpatient) did we complete? ☐ How many hours of treatment did we provide?	☐ What percent of service was in-home? ☐ How long is our waiting list for service? ☐ How long until the next opening in the appointment schedule?
Output	For our client population: ☐ How many clients are living safely at home? ☐ How many of our clients are in school or employed?	For our client population: ☐ What percent of our clients are living safely at home? ☐ What percent are in school or employed? ☐ What percent show demonstrable improvement in functioning?

Juvenile Justice

	Quantity	Quality
Input	☐ How many children are in custody (by age, offense and type of placement)?	☐ What percent of children are in community based vs. institutional care? ☐ What is the average caseload for juvenile probation workers? ☐ What percent of children in custody are in school or training?
Output	☐ How many children in custody are repeat offenders? ☐ How many showed an increase in the seriousness of offense? ☐ How many children who leave the system are in school or jobs?	☐ What percent of children in custody are repeat offenders? ☐ What percent showed an increase in the seriousness of offense? ☐ What percent of children who leave the system are in school or jobs?

Child Care Licensing

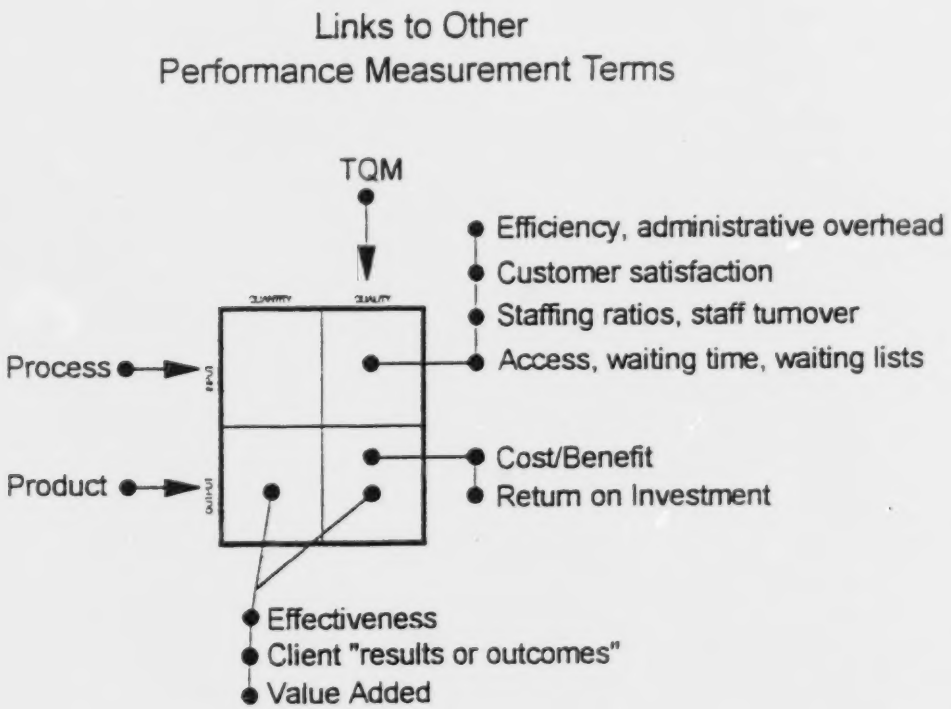
	Quantity	Quality
Input	☐ How many applications did we process? ☐ How many inspections did we do? ☐ How many recruitment sessions did we conduct?	☐ What percent of applications were processed on time? ☐ How many complaints did we receive about delays? What percent of total applications?
Output	☐ How many new centers, new child care slots were opened?	☐ What percent of licensed child care providers met safety/quality standards? ☐ How many child injuries in care were reported? ☐ What percent of child care is provided by licensed vs. unlicensed providers?

D. Links to Other Performance Measurement Frameworks

One of the interesting features of the four-quadrant framework is its ready connection to terms used in past work on performance measurement. This connection can help explain how other uses of terminology address different dimensions of performance measurement.

Consider, for example:

1. Efficiency and effectiveness: This is the classic set of terms in performance measurement, an age-old, time-honored, and generally usable approach. Efficiency measures are upper-right-quadrant measures that typically take the form of ratios of activity to resources. For example, cost per client served; direct service as a percent of total agency expenditures, or its inverse; administrative, overhead costs, percentage of total expenditures, are all measures of service efficiency. Such statements can not usually stand alone. A highly efficient service might not be a very good one. We must look to other quality-of-service (upper-right quadrant) measures like customer satisfaction. And measures of efficiency must be paired with the lower-left- and lower-right-quadrant-statements about what is produced—effectiveness (e.g., number or percentage of clients placed in jobs; number or percentage of students who graduate and go on to employment). The efficiency and effectiveness construct accounts for portions of 3 of the 4 quadrants.



2. Cost-benefit and return-on-investment measures are enormously important lower-right-quadrant measures of output. Cost-benefit ratios compare the quantity of benefit (lower left) to the cost of that benefit (for example, cost per job placement for an employment program, or cost per dollar collected in a child support enforcement program). This ratio goes beyond stating how much was produced and tells something about the quality of the production process itself, or how much we are getting for our expenditure. (Remember, the quadrants are measuring the service itself, as well as the service "products.") Taking cost-benefit measures a step further, we have rates of return on investment, which are also lower-right-quadrant measures. (By some estimates, for example, we reduce or avoid $10 to $14 dollars in health care costs for each dollar expended on immunizations.[12]) When this information is available—or can be created—for human service programs, it can be of great value in choosing where best to invest money to produce the optimal set of "client results."[13]

3. Customer satisfaction: Measures of customer satisfaction are permanent residents of the quality-of-service-delivery (upper right) quadrant. Such measures almost always capture important information about how well service is delivered. Customer satisfaction can tell us if the service is timely or accessible, or if the workers are courteous and helpful. It is possible, however, to have customers who are perfectly satisfied with a poor-quality service. We might find, for example, that drug-treatment clients in a poorly performing addiction treatment program are very satisfied with the service, in part because it does not push them very hard to change behavior. In this case, customer satisfaction does not measure the quality of output.

Customer satisfaction **can** be a measure of output quality in enterprises where products are sold to customers (e.g., cars) or where the **service is the product** (e.g., haircuts). In child-support enforcement, for example, customer satisfaction is probably a good (lower-right) measure of the quality of outputs, since single-parent customers who do not receive required child-support payments are not likely to be satisfied with the service. The fact that they may be treated courteously (upper-right measure) will not count for much in comparison.

The point is that the decision about where to place customer-satisfaction data, and how to interpret and use the data, **depends on the service itself**. The "right place" for each performance measure can be found by asking which of the four questions in the chart in Section III-A is answered with the help of these data.[14]

The chart in this section shows the link of other commonly-used performance measures to the four quadrants.

---

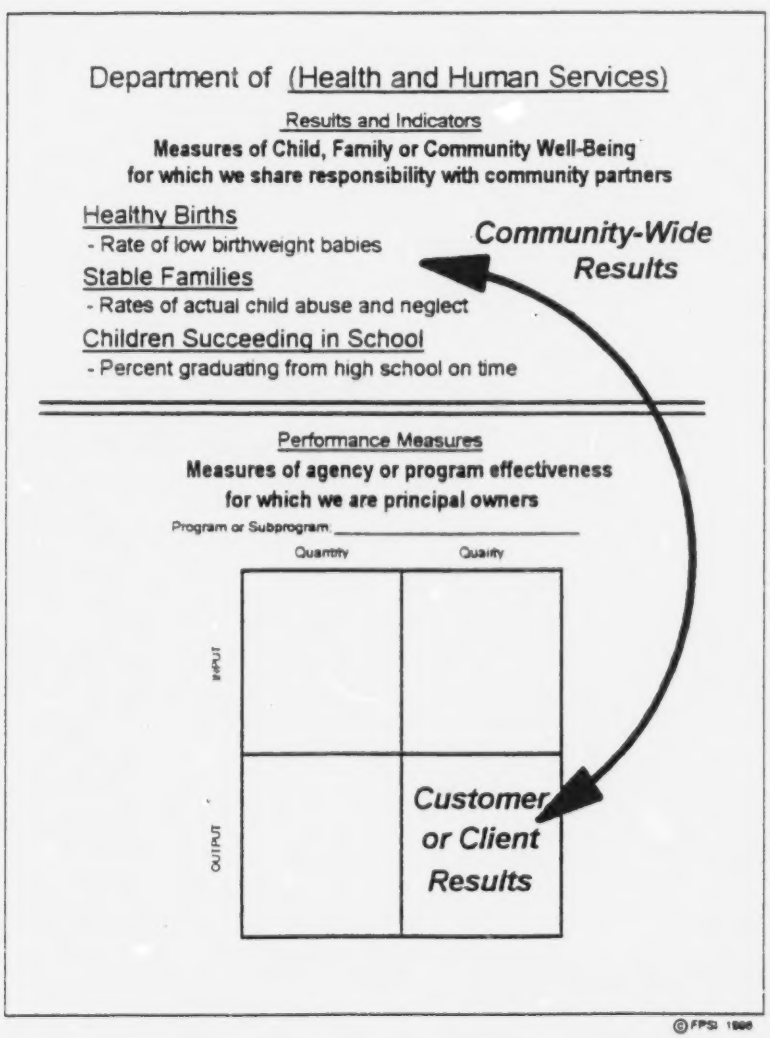
[12] "Report on Children Action Network," American Academy of Pediatrics News, 1991; as referenced in *Ready, Willing, and Able?*, The National Association of Child Advocates, 1996, page 27.

[13] See *Deciding for Investment, Getting Returns on Tax Dollars*, Alliance for Redesigning Government, National Academy of Public Administration, Jack Brizius and The Design Team, 1994. Iowa's Department of Management and the Iowa Council on Human Investment have made significant progress using this approach.

[14] This is an important lesson that has emerged from recent work using the four-quadrant approach. We do not need a complex set of rules about how to sort performance measures. Use of the questions as a sorting guide is a way to keep the common-sense nature of this work in focus.

E. The Link between Community-wide Results and Client "Results"

An interesting and important connection emerges when we examine performance measures from the quality/output quadrant. Many of these measures **sound like** indicators of well-being for children, families, and communities, exactly what we were measuring in the cross-agency results-accountability system. The relationship is shown on the following chart.


Department of (Health and Human Services)
Results and Indicators
Measures of Child, Family or Community Well-Being
for which we share responsibility with community partners
Healthy Births
- Rate of low birthweight babies
Stable Families
- Rates of actual child abuse and neglect
Children Succeeding in School
- Percent graduating from high school on time
Community-Wide Results
Performance Measures
Measures of agency or program effectiveness
for which we are principal owners
Program or Subprogram:
Quantity
Quality
INPUT
OUTPUT
Customer or Client Results
©FPSI 1996

This alignment of performance measures at the program level with results and indicators at the community level is a highly desirable characteristic of the four-quadrant framework, and is not surprising when you think about it. Often the difference between what we are trying to accomplish at the program and community levels differs only in matters of **scale**. Quality of output performance measures for programs will often be similar to community-wide indicators except for the **scale difference between a client population and the total population**. This may allow us to use the results/indicator framework adopted by a state, city, county, or community to "test" the selection of performance measures by various agencies to see if program-performance measurement aligns with what we are trying to accomplish at the community level.

A more important aspect of this alignment relates to the role of performance measures in funding what works. A central challenge in results-based budgeting is the development of **cross-agency and cross-community** strategies to measurably improve child and family well-being.[15] Lower-right-quadrant performance measures can help identify the best programs to include in such strategies, by showing how candidate programs improve, or fail to improve, the well-being of children and families in their client populations.

F. Results-based and Performance-based Budgeting Formats

These relationships also begin to suggest an approach to organizing and formatting a results-based budget. Such a format would incorporate both the broad cross-agency strategies to turn the curve on indicators of well-being, and also the detailed budgets for individual agencies and their programs. In effect, such a budget would have two sections, comparable to the top and bottom of the page in the chart above.

Part I of this budget would be organized **by result**, presenting strategies that cross agency and program lines. Several pages would be devoted to the presentation of the strategy for achieving a given result, such as "children succeeding in school." This presentation would include the following sections:

Section 1: Baselines: The history of our past performance on the three most important indicators of "children succeeding in school." Also, a presentation of our best forecast of where we are headed on these indicators if we stay on our current course. Usually, this involves a range of forecast scenarios (best case, likely case, and worst case).

Section 2: The story behind each of these baselines: Why do the baselines look the way they do? What got us to where we are now? What are the forces at work? What is our reasoning behind the forecasts?

Section 3: What Works: What does our experience tell us about what works in order to do better than the baseline? What does research tell us (if anything)? What has worked in other jurisdictions?

Section 4: Strategy: What have we done and what do we propose to do to improve? What is our cross-agency, cross-sector strategy to do this over the next several years.

Part II of the budget would present budget information not by result, but by agency, program, and sub-program. This is the way most budgets are now organized. This section of the budget would present performance measures in a way that parallels the use of indicators in Part I. So, for each agency, program, and sub-program, the document would present:

---

[15] See the discussion of "what works" in *A Strategy Map for Results-Based Budgeting*, pages 33-35.

Section 1: Three baselines with forecasts for the three primary performance measures for a given program or sub-program;

Section 2: A description of the story behind these performance trends;

Section 3: An explanation of "what works" to turn the curve toward improved performance; and

Section 4: A presentation of the action agenda and the budget to achieve improved performance.

This approach to displaying results and performance measures in budget documents will be addressed in more detail in future Finance Project papers on results and performance accountability.

IV. CHARACTERISTICS OF AN EFFECTIVE PERFORMANCE MEASUREMENT SYSTEM

Before we address how to implement a performance measurement system, let's review some of the characteristics of an effective system.[16] These characteristics are not prerequisites, but rather optimal qualities of a fully developed system, which may be useful in guiding the development process.

Six Characteristics of an Effective Performance Measurement System

Credible	**Fair**
Clear	**Practical**
Adaptable	**Connected**

A. Credible

The foremost requirement for a performance measurement system is **credibility**. Policy makers and citizens must have confidence that the information produced is accurate and relevant. Performance measures must be credible representations of the quantity and quality of the services provided by an agency or program.

Credibility is partly a matter of the objective accuracy of data, and partly a matter of the beholder of those data. Performance measurement systems must stand the test of capturing what is most important about a program's performance, both for those managing the program and those judging its performance. Performance measures that reflect only inputs or the quantity of goods and services provided by an agency will usually fall short on this criterion.

States such as Florida and Minnesota have bolstered the credibility of their indicator and performance measurement systems by documenting their data systems in considerable detail. Both states describe the reason each measure is important, what is being measured, and the data source. External review of performance measures by an independent body is

[16] If your tolerance for largely rhetorical stuff is low at the moment, you may want to skip ahead to Section V, and come back to this later.

another important strategy for making the data credible and powerful. In Texas, the State Auditor's Office reviews performance measures for accuracy. The Texas State Auditor's Office also issues guidelines for agencies about how to establish controls over data entry. State agencies in Texas must explain how they calculate performance measures, and retain documentation to support the calculations.

B. Fair

Performance systems should, to the greatest extent possible, provide **fair** gauges of agency and program performance. This means that measures should generally reflect factors and products that agency and program managers can influence or control. But there is an important qualification (perhaps trap) here. There is arguably **no** program effect that is **totally** within the control of program managers. Social programs operate in complex environments where performance is affected by economic, demographic, and other forces outside the program's control. This should not serve as an excuse to avoid performance measurement and accountability, but should help in both choosing and interpreting performance data. If control (fairness) were the overriding prerequisite for performance measures, then there would be no performance measures.

While no manager controls **all** the factors that affect program performance, it is legitimate for measurement systems to concentrate on bottom-line quality measures, and stretch people to think of ways in which they can partner with others to leverage resources that they **do not control**, in order to improve performance. Child welfare managers can partner with police and court officials to improve responses to child-abuse reports. Education managers can partner with health and human service providers to improve school achievement for children in troubled families. Juvenile justice officials can partner with community organizations to improve recidivism rates. Performance measurement can be used both to account for what people do with what they have **and** how well they collaborate with others who control resources vital to the program's success.

Fairness is as much a matter of how data are used as how they are selected. As discussed below, performance measurement should not be used as a blunt instrument to punish poor performance, but as a tool to improve performance. However, performance measures that attempt to hold public officials accountable for matters wholly beyond their control fail the fairness test (and will usually fail the utility and credibility tests as well). A common mistake by many states and communities is to use indicators and performance measures interchangeably, holding public agencies accountable for both. When performance measures appear to be unfair, they often turn out to be **indicators** of cross-agency or community well-being, rather than measures of program performance.

C. Clear

Performance measures should be **clear** and easy to understand and use. If performance measures are too complicated, they will be of little use in helping decision makers and citizens understand program performance or pointing out where improvements are needed. For example, decision makers and the public may be able to understand data on the percentage of juvenile offenders who commit additional crimes, but they will be much less able to understand or use a regression-based eight-part composite index that compares actual rates of recidivism to projected rates.

Often, it is not the performance measure itself that lacks clarity, but rather the way that the data are summarized and explained. If a school district reports that students who took Advanced Placement (AP) courses averaged a score of 3 on the AP exam, it is hard to interpret what this means. But if we are also informed that 40 percent of students taking the exam received college credit, then the performance measure is clearer and more useful for policy makers and ordinary citizens alike.

D. Practical

The performance measurement system should be **practical** to administer and implement. The way in which data are collected is a major factor in practicality. A good performance measurement system requires a significant and sustained investment in data collection. Since data collection is expensive (both in terms of dollars and agency-worker time), agencies must carefully weigh the value of performance measures, the investment in collection, and alternate ways to collect this data (e.g., 100% reporting vs. samples and surveys).

A well-defined data collection strategy is one that does not simply overlay worker functions with data collection requirements, but is built around line-workers' jobs in such a way that the data system becomes a tool to assist in performing those jobs. Performance measurement requirements for an agency should be a natural by-product of such a system. Consider the ways in which airline reservation systems have been designed to assist the poor soul trying to check in a 747-worth of passengers. The needs of company executives follow suit. Imagine trying to check in for a flight with a system designed primarily to meet executive needs.

Another dimension of practicality involves the development, operation, and linkage of data systems. Different agencies often collect information on the same people. While it is difficult to do, it makes sense for agencies to coordinate and, where possible, share data-collection strategies and instruments. Presentation of performance data at the county, city, and community levels also makes the information more useful. School system data on educational performance, for example, may be relevant to county or school system policy makers, but data on educational performance by school will more directly help principals and parents attempting to increase student learning. Data collection and analysis should support efforts to improve performance at all levels of the system.

E. Adaptable

As public goals and policies change, performance measurement systems must adapt to reflect these changes. When programs change, data requirements often change as well. And performance systems need to keep pace with these changes. However, changes in data collection create problems of comparability with prior-period data. And this requires an increased measure of analytic sophistication in tracking performance across discontinuities in policy.

The most important "adaptability" challenge may be the progressive development of less categorical cross-agency service systems (including managed care) for children and families. These changes hold real promise for more effective and more responsive services. But "less categorical" does not mean, and cannot mean, "less accountable." New cross-agency and cross-community service structures will create demands for improved tracking of service effects, even as the categories that underlie traditional reporting are phased out. Performance systems must develop in parallel with service-system development, so that we have and maintain the tools to manage and account for our performance.

F. Connected

Finally, performance measures must be **connected to and integrated with** other aspects of public planning, budgeting, and management systems. Performance measures are designed to provide feedback about the effectiveness of agencies, programs and policies. In order for that feedback to make a difference, it must be integrated into management systems (so that programs can be modified to perform better), budgeting systems (so dollars and other resources can be focused on programs that work), and accountability systems (so that managers can be rewarded for outstanding performance and helped to improve when

performance is poor). Some of the specifics of how to design and implement a performance measurement system that is integrated with planning, budgeting, management, and accountability systems are discussed in further detail in the next section.

V. USING PERFORMANCE MEASURES TO IMPROVE PERFORMANCE

The principal purpose of performance measurement is, not surprisingly, to improve performance. So far, we have dealt only with how to select data, and the principles of measurement systems, not how performance data might be used. In this section, we offer a few ideas about how to use performance measures.

A. Building a Performance Measurement System from the Bottom Up

Whatever else may be true of performance measurement systems, they almost always **display** too much, not too little, data.[17] Typically, for each sub-program, 10 or more performance measures are shown. As we move from sub-program to program to agency levels, the number of **displayed** performance measures grows exponentially. This provides executive and legislative branch decision makers with a sea of data,[18] but no particular way to sort out what is important from what is not.

While it makes sense to build performance measurement systems from the bottom up, this does not mean that we must adopt the undisciplined practice of using unlimited numbers of performance measures. The first and most important feature of a good performance measurement system is the use of a common-sense approach to "seeing the forest for the trees".

The first task is to contain the data explosion at each step in the construction process. For each level of performance, we identify the 2, 3, or 4 most important performance measures. Measures not selected here are still important, but do not need to be reported outside of that particular performance level. The four-quadrant sorting bin displayed previously can be used to help select these measures at each step in the process.

Using this approach, each level of the performance document or budget has **the same amount of performance information organized in roughly the same way**. Agency X monitors its performance on 3 to 4 primary measures. Program X monitors its performance on 3 to 4 primary measures. And so forth. More detail is found in each successive level.

In an agency with three levels (agency, program, sub-program) it works like this:

1. For each sub-program:
   - Identify the "candidate list" of performance measures available in the four quadrants above.
   - Pick the most important 2, 3, or 4 primary measures. These should generally come from the right-hand quality quadrants. (See Section C below for additional criteria for selecting primary measures.)
   - Create baselines with forecasts for these measures.
2. For each program, repeat this process, using the performance measures of the program's sub-programs as the candidate-measurement list.
3. For the agency as a whole, repeat this process using the agency's program-level performance measures as the candidate list.

---

[17] This does not mean that there is too much data from which to choose, only that too much of what is available is displayed.

[18] From 15th-century navigational charts, the exponential monster in the Sea of Data (Eighth of the Seven Seas).

In the course of this work, it is not uncommon to find programs, and even whole agencies, for which there is very little good data. In this case, the data selection process is not about picking the best of good data candidates, but finding **any** good data candidates. There are rarely any easy answers to this problem. But it is important—even with limited data—to proceed with development of performance measures and to improve the system over time. It is sometimes possible to create data, based on sampling techniques (by reading a limited number of case records, for example) as a short-term substitute for later data system development.

A related problem has to do with the **relative scarcity of quality measures** in data-system reports. Most agency data systems count quantity, not quality. Here, one relatively simple solution involves the use of "composite" performance measures, that is, performance measures that are created by calculating the ratio of two existing quantity measures. For example, many agencies count the number of safety or compliance violations among the programs they supervise. By itself, the raw count of violation totals does not mean much. But by calculating the ratio of program components with reported violations to total program components, a useful measure of quality can be created. Most good quality measures, whether currently reported or proposed, take the form of composite measures.

B. Building a Performance Measurement System from the Top Down

(or a word about that rare occasion when a top-down approach makes sense)

One of the most common mistakes in the use of performance measurement in management and budgeting is the tendency to implement performance measurement **all at once**, on a grand scale: "Starting next week, every manager of every program and sub-program must begin reporting on performance." Mountains of paper are produced. Little of it is used for anything. People come quickly to resent the intrusion of these new, time-consuming, and largely useless tasks. And the system is eventually abandoned.

There is nothing wrong with having performance measures for every component of an agency. But consider a different way of getting there. Imagine that the agency director asked each of the people who report directly to her or him to bring a few performance measures with them to their next meeting. This could take the form of the four-quadrant chart filled in with one entry in each quadrant. They could discuss three things:

- What do these data tell us about performance?
- What more would we like to know? (For example, comparison to last year, last month, 1-, 2-, or 5-year trends, maybe forecasts of performance.)
- Are these the right/best performance measures? The four-quadrant chart could be used to add or drop performance measures in these first meetings.

This process could, over a few months, lead to the creation of a regular performance report to be reviewed at each meeting. Over time, the performance measures could become the basis for agreeing on agency, or even personal, goals for performance (and, in the most advanced scenario, could be used for performance "contracting" between the agency head and the program manager).

By starting the process this way (or using this method to build on an existing performance measurement system), two very important messages are sent:

- Performance measurement is part of day-to-day management. It is not some back-burner, humorless, tedious, and irrelevant exercise; and
- Top management is modeling behavior for the rest of the organization.

This is why the top-down approach makes sense in this case. This allows, even encourages, the senior management to use this same process with the people who report to

them, and to **build down** through the organization. (This is not the way most management books tell you to do it, but it probably works better.)

Still another reason why working from the top down makes sense is that the performance measures of individual programs and sub-programs should be tied to the most important performance measures for the agency as a whole. If it is done right, working top/down will give people a sense of what top management sees as important, without making this an inflexible and domineering perspective.

The best work on performance measurement will be iterative, top/down **and** bottom/up. But top-down work of any sort has taken such a beating in the management literature that we sometimes don't recognize the times when it has a legitimate and important place. This is one of those times.

C. Selecting the Most Important Performance Measures

Primary vs. Secondary vs. Tertiary Measures

As we have seen, not all performance measures are created equal, and very few performance systems provide a disciplined focus on a small number of the most important measures. In the "Strategy Map" paper, we put forward the notion that not all **indicators** are equally important.

This same principle applies to performance measurement. We need a system in which each program (and each agency) is required to select the most important measures of performance and use these as the focus of performance reporting and accountability. These "primary" performance measures should be selected using the following criteria:

1. Measures should be given priority—as shown on the chart in Section II—as follows:

1st:	Quality of outputs
2nd:	Quality of inputs
3rd:	Quantity of outputs
4th:	Quantity of inputs

2. Primary measures should then meet the same three tests applied to indicators:[iv]

   - **Communication Power**: Does the performance measure communicate with both internal and external/public constituencies about "how we are doing"? It is possible to think of this in terms of a **public-square test**. If it were necessary to stand in a public square and explain the performance of your program with only two or three pieces of data, what data would you use? Obviously you could bring a thick report to the square and begin a long recitation, but the crowd would thin quickly. No one will listen to, absorb, or understand more than a few pieces of descriptive data. They must be powerful, common-sense, and compelling, not arcane and bureaucratic measures. The point here is to develop performance measures that have power and clarity with diverse audiences.

   - **Proxy Power**: Another simple truth about performance measures, like indicators, is that they tend to run in herds. If one is going in the right direction, chances are that many of the rest are as well. You do not need 20 performance measures telling you the same thing. Pick the ones that have the greatest proxy power (i.e., those which are most likely to match the direction of the other measures in the herd).

[iv] See the "Strategy Map" paper (pages 13-14) for a fuller discussion of these criteria and their related application to the selection of indicators.

- **Data Power**: And last, but not least, it is important that the performance measures we choose are ones for which we have quality data and which allow us to see progress—or the lack thereof—on a regular and frequent basis. Performance measures should be available on at least a monthly or quarterly basis. This allows managers and others to plot the new point on the curve and assess how we are doing in relation to the baseline.

Performance measures that are not selected as primary measures become part of the **secondary** list of performance measures that can be used in agency management and operations processes. The **tertiary list** consists of performance measures to be developed or improved. It includes the data agenda for future development.

You may have trouble getting people to limit the number of performance measures to 3 or 4. The discussion of the "credit trap" in the next section explains why people feel the need to see their particular performance data among the selected measures. In one county budget, it was very important for the Economic Development unit to communicate the **quantity** of the work that they had done: how many requests for information had been processed, how many businesses had been assisted, and how many publications had been distributed. Only after this information was presented did they get to the matter of outputs: how much new business was developed, "cost per job created," and "cost per dollar of nonresidential investment."

If you can't get to the 2, 3, or 4 most important quality-output measures, the next best thing may be to show all four types of performance measures, and use that display as the basis for discussing what is really important. Is it the number of information requests processed, or number of jobs created? If you are in the economic development business, it will be obvious that "jobs created" data are more important. A reporting or presentation format that uses this approach might have sections that separately present quantity and quality measures, and then go on to analyze the more important quality measures.

D. The Matter of Baselines

Defining performance success as "turning the curve"

We often set ourselves up for failure in our work on performance measurement by creating unrealistic expectations and impossible standards for success. A large part of this problem is attributable to defining success by "point to point" improvement:

> "Our rate of youth violating probation is x%. Success means decreasing this rate by 25% over the next 2 years."

Agency performance conditions, just like the indicators of child and family well-being, tend to be more complex than this. These conditions have direction and inertia. This is reflected in a baseline, which is sometimes headed in the wrong direction. These directions cannot always be changed quickly.

Sometimes the best we can do, in the short term, is to slow the rate at which things get worse before we can turn the curve in the right direction. This is a more realistic way of thinking about success (and failure). Success is turning away from the curve or beating the baseline, not turning on a dime to achieve some arbitrary lower target.

Each baseline, in turn, has two components: an historical component and a forecast component. Forecasting is at best an inexact science, and forecasts should reflect a reasonable range of possible future courses: high, medium, and low, or optimistic, best guess, and pessimistic. While forecasting can be difficult and even risky, the forecast component is very important. First, it communicates a powerful message about what we can expect to happen if we stay on our current course. It can be used to frame the fundamental question in this work: whether that expected course is an acceptable one. Second, it provides a reference against which to look at data as they come in and make judgments about how we are doing month to

month, quarter to quarter, and year to year. These kind of processes can and should be dynamic, using data to test ourselves and our strategies on a regular basis.[20]

This view is common in private-sector sales operations, where sales objectives are set—not in relation to last year's absolute level—but in relation to the "normal" expected growth in sales. Some of this growth derives from such forces as population growth and inflation. While the sales analogy may not translate cleanly to human services, the fact that similar "market" and demographic changes affect the likely future course of performance does translate. Child-support collections, for example, are affected by employment rates and changes in wages. Improving performance in this service should be geared to exceed **expected** changes related to these factors.

Baselines are therefore an essential component of performance measurement within results-based decision making and budgeting systems. Without baselines, we are blind to the reality of complex problems and complex performance environments. We are limited by systems that inaccurately measure progress and which skew decision making away from investments. Baselines allow us to think about problems in multi-year terms and avoid the oversimplifications that accompany year-to-year or point-to-point comparisons.

In one city budget, performance data for the fire department showed very favorable rates of fire incidence, injury rate, death rate, and property-loss rate compared to national averages. While these ratios are significant measures of fire department performance, a more important picture might be the **trends** over the last several months or years. If, for example, property loss rates **doubled** in the last year, this would constitute a serious performance problem easily masked by a favorable point-in-time comparison to national averages.

E. The Matter of Standards

Standards have an important place in work on performance measurement. And that place is with the two quality quadrants. We have a long history of developing and using standards to gauge quality—from child care staffing ratios to automobile gas-mileage standards. The four-quadrant approach provides a clear place to ground the use of existing standards and the development of new ones.[21]

Let's look at some examples of standards in each quadrant:

Quality of Service Delivery (Upper-Right Quadrant):

- **Timeliness**: Standards are often established for response to inquiry, decisions on applications, and, sometimes, waiting time for service. For example, child-welfare laws often require that the investigation of an abuse report be initiated within 24 hours. State and local agencies sometimes establish minimum performance standards for these rates in the 95% to 100% range.
- **Accessibility**: There are well-established standards with regard to handicap accessibility. Other accessibility standards have to do with office network coverage, convenience of public transportation, and hours of operations.

---

[20] There is a growing literature on self-evaluation. See "Improving Evaluability Through Self-Evaluation," Charles L. Usher, *Evaluation Practice*, Vol. 16, No. 1, 1995, pp. 59-68, or *Empowerment Evaluation: Knowledge and Tools for Self-assessment and Accountability*, D. M. Fetterman, S. Kaftarian, and A. Wandersman (1995), Thousand Oaks, CA: Sage.

[21] Note that standards are not performance measures, but desired values for performance measures. Standards are therefore not "entries" in the quality quadrants, but values associated with such entries.

- **Staffing ratios**: Among the best-known and sometimes most controversial standards are those established for the ratio of staff to clients/customers in various services. For example, child-care laws set standards by type of child care and by the age of the children in care. Similar standards often exist for group or institutional care for children in foster care.

Quality of Service Product (Lower-Right Quadrant):

Lower-right-quadrant standards are much more rare, and, in some cases, necessarily experimental.

- **Client condition standards**: These are standards that address rates of improvement/deterioration in client conditions (e.g., recovery rates for routine surgery at hospitals; juvenile justice escape or recidivism rates in privatized detention facilities; or job placement and retention standards under welfare-to-work programs).
- **Environmental standards**: Clean air and water standards for specific industries are lower-right-quadrant standards. These are quality-of-output measures. (They illustrate well how controversial lower-right standards can be.)

Standards for (upper right) service delivery are easier to define than (lower right) quality of client conditions achieved. For many services, we do not know enough about what level of quality/output performance is achievable to set standards. And different service systems often do not provide a level playing field to compare provider performance to a given set of standards. This does not mean that we should not move to test and eventually adopt such standards in both quadrants.

In the meantime, we have two usable substitutes for standards: the creation of **baselines** for prior performance and the use of **"benchmarking"**[22] against other similar programs and agencies. In the case of baselines, we can **test our performance against our past record** and try to do better than the baseline. This approach can serve many of the same management purposes as standards, and is a much more fair test of performance in the absence of good data on what is, in fact, achievable.

The **term "benchmark" is used in the private sector to describe a level of achievement of a (successful) competitor**. This is a powerful point of reference; and the performance levels of the most successful companies in a given industry often constitute a set of de facto standards for that industry. The counterpart in family and children's service programs is the comparison to performance in other states, counties, cities, and communities. When using these types of comparisons as a substitute for standards, it is important to consider differences in the socio-economic "operating" environments, just as industries (sometimes) adjust benchmarks/standards for differing market conditions.

[22] Note the difference in the way in which the term "benchmark" is used in public- and private-sector applications. The public sector often uses the term "benchmark" to mean an indicator or performance measure. The private sector uses the term to mean a particular level of (desired and achievable) performance. (And, yes, this is the same footnote that appeared in Section II.)

VI. EXAMPLES OF STATE AND LOCAL PERFORMANCE MEASUREMENT SYSTEMS

Looking at state and local performance measurement through the four-quadrant lens.

In this section, we examine several states that have well-developed performance measurement systems. The descriptions that follow are not intended as either critiques or full summaries of these states' systems, but rather a view of the state's framework through the lens of the approach presented in this paper. In each case, these states have put enormous thought and energy into the development of these systems, and each has many features worth studying and replicating. The bibliography provides references to the budget source documents from each state and to a number of reports summarizing these efforts. Note that these are generally examples of **performance measurement** systems, **not results-based** systems. As noted in previous sections, these are connected, but separate, areas in which to excel. For the best examples of results-based decision making and budgeting systems, see the papers referenced under that heading in the bibliography. Note that the following sections generally use terms and definitions as they appear in the states' documents, and language usage has not been edited to conform to the definitions offered in Section II.

A. Texas: Strategic Planning and Budgeting System

Texas has one of the most advanced performance measurement systems among state governments. Established by legislation in 1991, the State's Strategic Planning and Budgeting System has a four-part structure (Planning, Budgeting, Implementation, Evaluation) and six stated objectives, paraphrased as follows:

- Focus the appropriations process on outcomes.
- Strengthen the monitoring of budgets and performance.
- Establish standardized unit-cost measures.
- Simplify the budget process.
- Provide rewards and penalties for success and failure.
- Assure the accuracy of measurement data (using a review and certification process by the State Auditor's Office).

Texas's strategic planning framework is built around statements of mission, goal, priority goal, result, performance measures, and objectives. The system sets out "Workload vs. Performance" as competing approaches to budget development, with the state choosing to take the performance road. Use of these terms in Texas closely parallels the distinction in this paper between effort (workload) and effect (performance). The Texas system focuses primarily on performance measures of quality and thereby avoids the most common mistake of performance measurement systems, a preoccupation with how much is done, not how well it is done.

The system currently uses four principle **types** of performance measures.

- **Outcome measure**: a quantifiable indicator of the public benefits from a state entity's actions.
- **Output measure**: a quantifiable indicator of a state entity's goods or services produced.
- **Explanatory/Input measure**: an indicator that shows the resources used to produce services or a factor that affects agency performance.
- **Efficiency measure**: a quantified indicator of productivity expressed in unit costs, units of time or other ratio-based unit.

These terms describe the types of measurement used in an over-arching system that moves in a structured process:

- **From Statements of Purpose**: statewide vision, mission, philosophy, functional goals and benchmarks and agency mission and philosophy,
- **to Statements of Direction**: agency goals, objectives, strategies, and action plans,
- **to Statements of Impact**: including outcome measures, output measures, efficiency measures, and explanatory measures.

State agencies set five-year goals through a strategic planning process and establish unit-cost measures for important activities. As part of their budget requests, agencies list each goal and the objectives and the strategies associated with it, along with the budgetary resources needed to achieve each goal. Agencies also list the performance measures associated with each goal, along with an estimate for the coming year. Agencies rank their activities and the funding needed for those activities in descending-priority order.

Every state agency is linked by computer to the Legislative Budget Board through the Automated Budget and Evaluation System, which integrates planning and policy goals, funding sources, spending-line items, and performance measures—including the definition of each measure, targeted and actual performance, and explanations of any variances. Agency performance is reported on a quarterly and an annual basis to the state auditor's office, the Legislative Budget Board, and the Governor's Office of Budget and Planning.

B. Arizona: Program Authorization Reviews

Arizona's budget offers an excellent example of the use of performance measurement to improve the performance of state programs, and the overall management of the state budget. Arizona makes use of scheduled Program Authorization Reviews (PAR's), which systematically assess program performance. This system links strategic planning, performance measures, program evaluation, and budgeting.

The Fiscal Year (FY) 1998 and 1999 Executive budget includes key performance measures for each budget unit. Agencies were instructed to provide a **one-page or less** summary of their most important performance measures with their budget requests. Those key performance measures were published as submitted by the agencies without modification.

During the FY 1998-1999 budgeting cycle, the (1997) Program Authorization Review process also reviewed the performance of 30 selected programs and sub-programs in 14 different state agencies. The PAR process addressed four key questions:

- How does the mission (of the program) fit with the Agency's mission and program's enabling authority?
- Does the program meet its mission and goals efficiently and effectively, including comparison with other jurisdictions?
- Do the program's performance measures and performance targets adequately capture these results?
- Are there other cost-effective alternative methods of accomplishing the program's mission?

One of the 1996 PAR findings, referenced in the FY 1997 budget, addresses the Department of Corrections and gives some insight into the type of performance measures being used. For each of the prison complexes, the following information was requested on a quarterly basis. These measures principally address how much service was delivered and how well service was delivered. The Department reported on:

- Average daily population
- Cost per inmate

- Percentage of corrections policies complied with
- Ratio of administrative to institutional staff
- Escapes per 1,000 inmates

The Executive Summary of the 1997 Program Authorization Reviews identified a number of key issues and conclusions from the second year of the PAR process, including the importance of customer-satisfaction measurement; the need to develop better historical data; the need to "benchmark" agency performance with other similar organizations; and the wide variation in the quality of agency self-assessment.

Two other conclusions bear on the matter of program and agency control of performance. The reviewers recognized the fact that many different programs within an agency contribute to the achievement of a particular agency mission. And some measures involve factors beyond the agency's control. In particular, the use of recidivism to measure the performance of the corrections system is complicated by the fact that such rates are dependent upon many outside factors.[23] In spite of these difficulties, the PAR reviewers, to their credit, reasserted their belief that "recidivism is a key measure in evaluating overall program performance" even though they recognize the agency cannot be held solely responsible for this result. And while the Department cannot fully control recidivism, they do provide opportunities for inmate rehabilitation.

Arizona offers an example of how performance measurement can be used in budget decision making without necessarily overloading the budget document with data.[24] And experience to date suggests that the PAR process is an effective way to use program performance data in order to improve program performance.

C. <u>North Carolina:</u> Performance Budgeting System

Beginning in 1991, North Carolina implemented a budgetary process that focuses upon performance measures for all state agencies. The process involves developing and using program performance measures for the more than 3,000 state-funded "activities" included in the departmental budgets. The North Carolina General Assembly has supported the movement away from a traditional line-item "input-focused" budget to an "outcome-focused"[25] analysis of how state dollars are expended, and the effects of such spending on the well-being of the state and the citizens served by the state's programs. This performance perspective is reflected in the FY 1997-1998 and FY 1998-1999 budget recommendations that employ a format designed to demonstrate the relationships and ultimate effects of similar services funded at the state level.[26]

North Carolina's Performance/Program Budgeting (P/PB) system covers all state and federal funded activities and allows for a complete classification for every component of state

---

[23] See the discussion of the alignment of client results and community-wide results (Section III-E) and the discussion of credibility and control issues (Section IV-B).

[24] Arizona uses an interesting "safety valve" for the data-overload problem, with the publication of a "Master List of State Government Programs," which provides a "more comprehensive listing of performance measures for every agency, program, and sub-program."

[25] Note that North Carolina uses the term "outcome" to describe the effectiveness of state programs. This generally corresponds to the use of "output" measures as described in Section III.

[26] These budget documents are available on the North Carolina Office of State Planning's electronic world wide web site at http://ospl.state.nc.us.

government. This approach to budgeting involves grouping government services that share a common purpose, have common clientele, or common programmatic outcome measures. There are ten "program areas": general government, human services, corrections, justice and public safety, environment, health services, transportation, education, commerce, and cultural resources. Funds for these areas are grouped together regardless of where they fit within the organizational structure. This budgeting approach is particularly useful for identifying instances where similar services are administered by different parts of state government and how these efforts could be better coordinated across organizational boundaries.

For accounting purposes, each "element" of state government is assigned a four digit code which allows ready identification of its alignment by program area, program and sub-program. For example, "elements" of the **Ground Water Quality** "sub-program," fit within the **Preserve and Enhance Water Quality** "program," which fits within the **Environment** "program area."[27] Another important feature of the North Carolina system is its use of multi-year baselines for key performance measures. In the Ground Water Quality section, for example, the budget presents a 10 year history of the number of contaminated wells by source of contamination.

North Carolina's Performance/Program Budgeting integrates planning, budgeting, and evaluation decisions by agencies. By linking measurable objectives to specific agency expenditures and performance measures, the consequences of budgetary decisions are made more explicit. This provides agencies, the legislature, and the general public with a better understanding of what can and should be accomplished by a particular level of program funding.

D. Other Notable State and Local Performance Measurement Systems

Many other states and communities have developed systems that link performance data to planning and budgeting decisions. While it is impossible to fully summarize or give credit to all such state and local efforts, some noteworthy systems are referenced below.

Communities such as **St. Petersburg Beach, Florida**, and **Phoenix, Arizona,** report performance data monthly and annually in order to compare performance to targets; **Phoenix** also conducts a customer-satisfaction survey every two years. **Indianapolis, Indiana** conducts regular citizen surveys; publishes a public budget document explaining resource allocations, as well as departmental goals and accomplishments in clear language; and uses performance data as a factor in determining pay increases. **Virginia Beach, Virginia** uses a performance measurement system as part of a Total Quality Management initiative covering all city programs. **San Mateo County, California** provides one of the best examples of performance-trend information in budget documentation and decision making. Each major program in the budget presents performance-trend data for two or three of the most important performance measures. The selection and presentation of these data make the information more relevant and useful in the budget decision-making process. **Iowa** is using performance measures to estimate the benefits from state expenditures, compare rates of return on program investments, and use these data in the review of agency budgets. In **Milwaukee, Wisconsin**, city departments must justify their annual funding requests in the context of a city-wide strategic plan, identifying objectives for the year 2000. Departments

[27] This coding structure is useful for planning within program areas, but does not fully address planning for results across program areas. The state's strategy for "children ready for school," for example, draws on sub-programs in the education, human services, and health program areas. See the discussion of the link between results and performance budgeting in Sections III-E and III-F.

must specify the dollars allocated for important activities and the impacts that they will have on Milwaukee residents. This process of activity-based costing, combined with performance measures, gives policy makers more comparable information about the costs of different services.

VII. LESSONS AND ISSUES

Performance measurement has a long history and a short memory. A number of lessons and issues from this work are summarized below.

A. The Language Trap

The language trap is the most common problem in building performance-accountability systems. Words matter. And our ability to communicate about complex subjects—such as agency, program, and community accountability—requires that we adopt understandable language conventions.

The discussion of "results, indicators, and performance measures" in Section II provides some essential distinctions for this work—most importantly, the distinction between cross-community accountability for results, and agency/program accountability for performance. When the vocabulary that we use fails to provide a ready means to keep this distinction clear, it is easy to confuse these two concepts and end up wrapped around an axle.

Language conventions should also be as simple and easy to understand as possible. **It should not be necessary to become an expert in the language of a performance measurement system in order to use it.** Frequent references to a glossary, or frequent debates about whether a measure fits into one category or another, may be signs that language conventions—and possibly the underlying framework—are too complicated.

B. The Bookshelf Trap

After language problems, the bookshelf trap is the next most common trap in performance measurement. Thousands of person hours may go into the production of a multi-volume performance measurement data set, which, when finished, is placed on the bookshelf and never used. This is the experience of more than one jurisdiction where performance documents have had limited utility in the executive branch budget process and are hardly used at all by the legislative branch.

What causes the bookshelf trap and how can it be avoided? First, it is important to remember that the challenge in this work is not to produce more paper, but useful paper. A one-page decision document that gets used is better than a hundred-page review of program performance that does not. The problem, of course, is that government is made up of layer after layer of organizational components. Any document that attempts to show even a small number of performance measures for all these levels would necessarily be long and complex.

There are several ways to avoid the bookshelf trap:

- **Make sure that the performance measures chosen are the ones used in day-to-day management of programs**. This means that the program director and agency director must share responsibility for choosing these measures. And they must be part of their management relationship. The test of "utility" will force the number into the reasonable range.
- **Keep the list of performance measures per program short**. And, of course, pick the most important measures - presumably from the quality column discussed above.
- **At each level of summation, drop all but the top three measures** from the organizational level below. At the agency level, the performance measurement (budget) document would show only three measures drawn from the program

measures below. At the program level, the document would show only three measures drawn from the sub-program level below, and so forth. If this discipline is too tough, then consider it an ideal toward which to aim. The point is to see the forest for the trees and to focus on the most important measures. This means that more detail is available **on request**. But not all detail is provided to all of the people all of the time.

Imagine that we are reporting to the stockholders of a large corporation. What do they need to know—bottom line—about the health and performance of this organization? This way of thinking about data presentation may help focus and simplify the work and avoid the bookshelf trap.

C. The Credit Trap

*(or, why people are so insistent on **quantity** measures)*

People want credit for what they do. They want management to understand how tough their jobs are. If they do more of something, they want it recognized. If the programs in which they work are understaffed, then they want management to acknowledge it and maybe do something about it. In this scheme of things, the overriding interest is in **how much** was done.

These are very common and understandable ways for people to think about and react to performance measurement. This reaction only becomes a trap if this view comes to dominate the process of developing performance systems and halts or overwhelms the development and use of other types of performance measures.

There are several ways to deal with this.

- **First**, performance systems and budgets should have a place for information about "how much was done"—the quantity of service delivered. It is useful to know if caseloads are going up or down. And programs should have a place to describe how many people they serve, as well as the size of the problems they face and the programs they administer. In some cases, this is important information in its own right, as in the case of the number of children enrolled in the school system or the number of families with children served by shelters for the homeless.

- **Second**, it is important not to stop there. Programs should be expected to go beyond "how much service was delivered" to "how well service was delivered." Caseload ratios, timeliness of service, and customer satisfaction are legitimate measures among others.

- **Third**, and most importantly, don't allow inputs (quantity or quality) to become the primary measures of agency or program performance. Program managers may be comfortable stopping here. Do not let this view prevail.

Remember that, in the industrial model, the number of client cases processed is an output. We now know that this is not an output in a change-agent-model service environment. We must look for real outputs for services delivered. And we must look beyond how much output is produced to how good these outputs are. In other words, let program managers get credit for how much they do. But force the issue of what is produced, and how well, beyond the old stock answers of cases, clients, and people served.

The credit trap is actually one of the causes of the bookshelf trap. If the purpose of the performance measurement system **becomes** "giving credit for work done," then the document will become a monster. People need to see their work measured and recognized in a public way. There is nothing wrong with this. But this means that organizations need other ways to recognize performance beyond the performance measurement system itself. Other **reward and recognition** methods are needed as part of the management mix. This makes it easier to resist the temptation to load so much onto the performance system that it lists

seriously to one side and eventually capsizes, sending all the little performance measures and their adherents to the bottom.

D. Pay for Performance

(and other matters of consequence)

Pay for performance is a growing practice of corporate America. A recent survey of 694 firms with over 5 million employees found that "29 percent of those firms are now using [pay for performance] types of incentive pay plans for hourly workers and non-management professionals...about three times what it was a decade ago."[28] At the executive level, the practice is even more widespread. In a 1992 survey by the Department of the Treasury, 38 of 41 "major American corporations" responding "use measures to link senior management appraisal and compensation to organizational performance."[29]

There are differing views about the extent to which individual or contract performance in services, and in particular public services, should be tied to rewards and punishments. The views range from the benevolent to the Machiavellian. The test should be the same, simple "what works" test used in the results-accountability framework. Does pay for performance work to improve performance? This may vary somewhat from place to place. But some simple principles seem to carry over.

- **There should be consequences for both good and bad performance.** Most job-satisfaction surveys show that money is not at the top of the list for job satisfaction. A sense of accomplishment and recognition is. Once you get past the survival and growth imperatives that go with organizations, the same is true.
- **Consequences should advance the overall performance of the organization.** This means that rewards and penalties should tie individual and unit behavior to the good of the enterprise. A performance-reward system used by Mobil Oil[30] weights the company's overall financial performance, the performance of a particular business unit, and the performance of the individual's work team. (Note that this did not include the **individual's** performance.)
- **People (and organizations) need to be treated with respect.** Most people (and most organizations) want to do a good job, and systems geared to treat everyone as if they are suspects in a job-performance scam will harm morale and performance.

Crafting money consequences to go with performance is a tricky business. Pay for performance is an appealing concept, but hard to implement when the products are changes in human conditions; when performance is often tied to the severity of client problems, not the quality of service delivery; and when there are often ready means to game the system. **This means that we should not rush to implement pay for performance (or other rewards and penalty policies) before we know what good performance is.** We need to build performance histories, and begin to measure and reward improvements on past performance.

---

[28] The survey by Watson and Wyatt & Co. was reported in *The Washington Post*, November 21, 1996, page D1.

[29] *Performance Measurement: Report on a Survey of Private Sector Performance Measures*, Department of the Treasury, Financial Management Service, January 1993, page 10.

[30] *The Washington Post*, November 21, 1996.

Over time, we can create performance standards that are fair and achievable. If we build systems and standards that people (individuals and contractors) consider fair, then we have a chance to improve performance, while minimizing gamesmanship.

E. Performance Anxiety

(and the link to Organizational Development)

Agencies are organic entities.[31] And how data are developed, distributed, and used are organic systems. Such systems can be healthy, or not. As a result, there is an important link between performance measurement and organizational development.

For many people, the only experience they have with performance measurement involves punishment. Data are used to distribute blame and, in some cases, pink slips. Why would anyone voluntarily produce data to feed this kind of monster? When organizations are operating in the blame mode, or feel that they are under siege, then the natural response is to make data-based accountability difficult. This is not difficult to do. When data are hard to get and of poor quality, it is hard to blame anyone for poor performance and still harder to prove the blame is deserved. Problems with data can always serve as the first line of defense. In business, this would be a formula for bankruptcy. In government, it can be standard practice.

How to get past the blame game is beyond the scope of this paper. But leadership and organizational development have a lot to do with it. The best-designed performance measurement system in the world will not work in a sick organization. And trying to put an ambitious system in place in such an environment will create resistance (often in the form of passive-aggressive behavior) and simply won't work. The necessary ingredient here is trust, and specifically, trust in the (reasonably) fair use of information. This is easy to say, but hard, even in good organizations, to practice.

If your organization is in this kind of "performance measurement equals punishment" trouble, or if you are operating in a hostile environment, then you need to be deliberate and strategic about putting such systems in place. And you need to think about how to use organizational-development "technology" to improve communication, trust, morale, and the other characteristics of successful and healthy organizations. While you are waiting for the ambulance to arrive, consider building performance measurement systems in a way that does not make matters worse. Rather than ask for a visible monthly performance report, develop the report in a private and confidential way until all players are reasonably comfortable with the data and how such data are being used. And work with the public-relations professionals in your organization to think about how to portray the good news, not just the bad news, that your organization produces.

Performance measurement is not a clean mechanical process. It is messy. It touches things that are important and it will generate strong reactions. The view that "information is power" is not an idle cliche. But if leaders and managers are in fact committed to doing better, then performance measurement is part of getting there. Think about how this part fits with other organic parts of your organization or system.

F. Auditing the Performance of Performance Measurement Systems

Looking at the many different performance systems now in use in state and local budget and strategic planning systems, one recurring impression is that they are very uneven in their implementation. One department does a good job identifying and reporting on performance measures. Another misinterprets the instructions and produces mush. Even where instructions are clear, there is wide variation in the quality of the work.

---

[31] Say that fast five times.

It may make sense to "audit" the performance measurement system. Such an audit would address both problems of implementation and problems in the system itself. By assessing the performance measurement system using the four-quadrant lens, it may be possible to see where the system could be improved. It would identify where individual departments could do a better job of identifying **and using** strong measures of performance. And it could also help define the link to results and indicators.

Texas has established an important role for its audit agency in verifying and certifying the performance measures used by state agencies. The Multnomah County, Oregon, auditor checks on how agencies are doing, and how the city is doing, in relation to its benchmarks—a role well above and beyond the traditional function of auditors. These may be models of what auditing agencies can become in the future when they move beyond narrow roles.

G. The Sole Ownership Trap

The distinction between results accountability and performance accountability helps explain one of the classic difficulties in budget reform efforts: the inability to make a one-to-one correspondence between results and departments. Most past approaches to budget reform put forward an uneasy compromise. Safety clearly depends on more than an effective police department, but we list all safety indicators only in the police budget. Success of children in school clearly depends on more than an effective school system, but we list education indicators only in the education department.

This need to have a single straight-line progression from result to department to program to performance measure is the hobgoblin of these reform efforts. There is a better answer. No department is, can be, or should be the **sole owner** of any **result**. Measuring success on results and measuring success on performance are **two different** (though interrelated) things. Departments can be principal owners, but they are not ever **sole** owners. This sounds like common sense. But it is rarely, if ever, seen in practice. People have been trying to reinvent the "straight progression" system for the last 50 years. This is a failure. It doesn't and can't work.

It may be neat accounting, but it is a poor representation of the way the world, let alone government, works. In results-based budgeting, each program can relate to as many results and indicators as make sense. It would be rare to find a program or sub-program that did not have multiple roles to play. Results-based budgeting allows these relationships to be used **in addition** to functional categorizations provided by traditional agency program descriptions.

H. Buyer Beware

Beware of reports (or consultants) who tell you that they have **the answer** about how to do performance measurement. Leaders in government need to be good consumers of advice, whether about performance measurement or anything else. This means looking at lots of models before you drive one home. The problem, of course, is that leaders are very busy people. So there is a temptation to take the first model that seems to work and leave it at that.

Buying advice is like buying a car. Look under the hood. Kick the tires. Take the time to compare models. Many in government feel that they do not have the time to be good consumers when it comes to planning and budgeting frameworks. This is not true, of course, with other forms of procurement, where we obsessively require competition against predetermined specifications. This paper is one of many to consider. The bibliography is a partial list of other documents and other approaches to read and consider.

You do not have to become an authority before you can choose. But the same principles of buying anything else apply: What do you need? How well does the approach that is offered meet your needs? We have all seen the frustration that comes with lengthy planning processes that are all process and no result. Taking the time at the beginning to chart a sound

course is the best answer. Take the best of what different people have to offer and then craft an approach that best meets your needs.

I. The Myth of Sisyphus

For some reason, almost all performance measurement systems use circular charts to depict the planning process. Just when you think you're finished, the damn thing starts over again. Here, we want people to have a sense of accomplishment and ownership, and what we give them instead is a version of hell of literally mythic proportions. We want to promote the idea of continuous improvement in the use of performance measures—and all of these processes are necessarily iterative—but, as you translate this work into your own environment, think of the poor soul rolling the performance measurement forms up the hill one more time, and find something other than circular imagery to describe the work.

VIII. CONCLUSION

Accountability systems—whether results or performance—are not ends in themselves, but means to the ends of improved conditions of well-being for children, families, and communities. The technology of accountability will always be developmental and controversial. If accountability is real, then it affects things that matter. It provides consequences for success and failure. Without such systems, we will fuel cynicism about government and private-sector performance, and, worse, we will deserve such cynicism. Performance measurement, as part of a results-based accountability system, can help build public confidence in government and community institutions, and, more importantly, help us create improved results for children, families, and communities.

BIBLIOGRAPHY

Following is a partial bibliography of results and performance accountability documents. Note that the use of language differs dramatically from document to document. And in most cases, the terms "result" and "outcome" are used to describe what are defined in this paper as agency or program "performance measures." Items that could be listed in more than one section are, generally, listed in the primary section to which the work applies.

Performance Measurement Guides

The Baldridge Award for Education: How to Measure and Document Quality Improvement, Jerome S. Arcaro, St. Lucie Press, 1995.

The Five Most Important Questions You Will Ever Ask About Your Nonprofit Organization, Peter F. Drucker, Jossey-Bass Publishers, 1993.

A Foundation for Success: A Guide to Performance Improvement for Literacy, National Institute for Literacy, Draft, June 1995.

Guide to Performance Measurement: State Agencies, Universities, Health-Related Institutions, Texas State Auditor's Office, Legislative Budget Board, and Governor's Office of Budget and Planning, August 1995.

Managing for Results and Measuring Success: Outcome Based Management for Human Services, Frederick K. Richmond and Eleanor W. Hunnemann, 1996.

Measuring Program Outcomes: A Practical Approach, United Way of America, 1996.

Outcome Funding, A New Approach to Targeted Grantmaking, Harold S. Williams, Arthur Y. Webb, and William J. Philips, Second Edition, The Rensselaerville Institute, 1993.

Performance Measurement Guide, Department of the Treasury, Financial Management Service, November 1993.

Who Will Bell the Cat? A Fable for Our Time: A Guide to Performance Measurement in Government, Price Waterhouse, 1993.

Reviews of Performance Measurement Practice

Improving Mission Performance through Strategic Information Management and Technology: Learning from Leading Organizations, General Accounting Office, May 1994.

Management Reforms, Examples of Public and Private Innovations to Improve Service Delivery, General Accounting Office, February 1994.

Measuring Performance in Human Service Systems, James F. Budde, American Management Association, 1979.

Monitoring the Outcomes of Economic Development Programs, Harry P. Hatry, Mark Fall, Thomas O. Singer, and Blaine E. Liner, Urban Institute Press, 1990.

Monitoring the Outcomes of Social Services, Annie Millar, Harry Hatry, and Margo Koss, Urban Institute Paper on State and Local Government, May 1977.

Performance Budgeting: State Experiences and Implications for the Federal Government, U.S. General Accounting Office, February 1993.

Performance Indicators in the Public Sector, Paul Jowett and Margaret Rothwell, MacMillan Press, 1988.

Performance Measurement: An Important Tool in Managing for Results, U.S. General Accounting Office, May 1992.

Performance Measurement in Selected Public Health Programs: 1995-1996 Regional Meetings, Department of Health and Human Services, Public Health Service.

Performance Measurement Lessons Learned, Thomas J. Cook, Jerry VanSant, Leslie Stewart, and Jamie Adrian, Research Triangle Institute, May 10, 1993.

Performance Measurement: Report on a Survey of Private Sector Performance Measures, Department of the Treasury, Financial Management Service, January 1993.

Program Performance Measures, Federal Agency Collection and Use of Performance Data, General Accounting Office, May 1992.

A Review of the Oregon and Texas Experience in Building Performance Measurement and Reporting Systems, Data Selection, Collection and Reporting, National Institute for Literacy, January 1995.

Fourth National Roundtable on Outcome Measures in Child Welfare Services: Summary of Proceedings, American Humane Association, Children's Division, 1997.

Toward Useful Performance Measurement: Lessons Learned from Initial Pilot Performance Plans Prepared Under the Government Performance and Results Act, National Academy of Public Administration, November 1994.

Using Performance Measures in the Federal Budget Process, Congressional Budget Office, July 1993.

Other Performance Measurement Reports and Documents

Deciding for Investment: Getting Returns on Tax Dollars, Jack Brizius and The Design Team, Alliance for Redesigning Government, National Academy of Public Administration, 1994.

Improving Government Performance: Evaluation Strategies for Strengthening Public Agencies and Programs, Joseph S. Wholey, Kathryn E. Newcomer, and Associates, Jossey-Bass, 1989.

Management Control in Nonprofit Organizations, Robert N. Anthony and David W. Young, and Richard D. Irwin, Inc., 1994. (See Chapter 12: Measurement of Output.)

Performance Measurement: The Key to Accelerating Organizational Improvement, Price Waterhouse, 1993.

Performance Measures for the Criminal Justice System, Discussion Papers from the BJS Princeton Project, U.S. Department of Justice, October 1993. (See "Measuring Performance When There is no Bottom Line" by John J. Dilulio, Jr., pp. 143-156)

Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector, David Osborne and Ted Gaebler, Addison-Wesley, 1992. (See "The Power of Performance Measurement," pp. 146-165; and Appendix B: The Art of Performance Measurement.)

Reports on Results Accountability

Aiming for Results: A Guide to Georgia's Benchmarks for Children and Families, Georgia Policy Council for Children and Families, 1996.

From Outcomes to Budgets: An Approach to Outcome (or Result) Based Budgeting for Family and Children's Services, Mark Friedman, Center for the Study of Social Policy, July 1995.

The Guide to Results-Based Accountability: An Annotated Bibliography of Publications, Web Sites and Other Resources, Harvard Family Resource Project, June 1996.

Making a Difference: Moving to Outcome-Based Accountability for Comprehensive Service Reforms, Nancy Young, Sid Gardner, Soraya Coley, Lisbeth Schorr, and Charles Bruner, National Center for Service Integration, 1994.

Oregon Benchmarks, Standards for Measuring Statewide Progress and Institutional Performance, Report to the 1995 Legislature, Oregon Progress Board, December 1994.

A Strategy Map for Results-Based Budgeting: Moving From Theory to Practice, Mark Friedman, The Finance Project, October 1996.

State Budget and Planning Documents

The following budget documents are referenced in the text:

Guide to Performance Measurement: State Agencies, Universities, Health-Related Institutions, Texas State Auditor's Office, Legislative Budget Board, and Governor's Office of Budget and Planning, August 1995.

Legislative Appropriations Request for the Biennium Beginning September 1, 1997: Detailed Instructions for Executive and Administrative Agencies, The State of Texas: Governor's Office of Budget and Planning and Legislative Budget Board, April 1996.

The 1993-95 Biennium: Primer to Performance Budget, The North Carolina State Budget, North Carolina Office of State Planning and Office of State Budget and Management, 1992.

Strategic Planning and Performance Measurement Handbook: Managing for Results, Office of Strategic Planning and Budgeting, Office of the Governor, The State of Arizona, May 1995.

The North Carolina State Budget, 1995-1997, Performance Program Budgets, North Carolina Offices of State Budget and Management and State Planning, December 1994.

About The Finance Project

The Finance Project is a national initiative to improve the effectiveness, efficiency, and equity of public- and private-sector financing for education, other children's services, and community building and development. With leadership and support from a consortium of private foundations, The Finance Project was established in 1994 as an independent, non-profit organization. It undertakes an ambitious array of policy research and development activities, policy maker forums and public education activities, and support and technical assistance activities.

The work of The Finance Project is aimed at increasing knowledge and strengthening the capability of communities, states, the federal government, non-governmental entities, and the private sector to implement promising strategies for generating necessary fiscal resources and improving the return on investments in children and their families. Its activities are intended to:

- Examine the ways in which governments at all levels, and the private sector, finance education and other supports and services for children (age 0-18) and their families;
- Identify and highlight structural and regulatory barriers that impede the effectiveness of programs, institutions, and services, as well as other public investments, aimed at promoting children's growth and development;
- Outline the characteristics of financing strategies and related structural and administrative arrangements that support improvements in education, other children's services, and community building and development;
- Identify promising approaches for implementing these financing strategies at the federal, state, and local levels and assess their costs, benefits, and feasibility;
- Highlight the necessary steps and cost requirements of converting to new financing strategies; and
- Strengthen intellectual, technical, and political capability to initiate major long-term reform and restructuring of financing systems, as well as interim steps to overcome inefficiencies and inequities within current systems.

The Finance Project's work is organized within three broad areas. Activities in each area build upon and inform the others:

I. Review and analysis of federal, state, and local financing strategies for education, other children's services, and community building and development.

Through a series of working papers and other studies, The Finance Project examines key issues and trends in federal, state, and local fiscal capacity and public expenditures on behalf of children and their families and assesses strategies and challenges for generating fiscal resources, planning and budgeting public expenditures, and measuring the impact of public investments. It also examines the implications of relevant statutes, regulatory provisions, and judicial decisions for patterns of expenditure. Approximately 40 studies have been conducted and published to date.

II. Development of ideas, options, and policy tools.

The Finance Project organizes and convenes interdisciplinary working groups to serve as development and design teams for policy-relevant ideas, options, and tools for financing reforms. Additionally, it initiates and participates in collaborative efforts to mobilize the intellectual and technical resources needed to bring about positive change.

The working groups bring together individuals and ideas from many domains to identify and assess alternative approaches for generating necessary fiscal resources and improving the return on public investments; to develop options for more effective fiscal policies and streamlined administrative practices; and to design prototypes, models, tools, and materials to support their implementation in the years ahead. Three working groups are developing tools and options in the following areas:

- Strategies for Generating Revenue for Education and Other Children's Services;
- Results-based Planning, Budgeting, and Accountability Systems; and
- Financing Comprehensive, Community-based Support Systems.

Additional working groups formed as needed.

Led by the Finance Project, the Collaborative Initiative to Improve Financing for Young Children and Their Families, a group of national organizations, as well as selected state and local leaders who are involved in efforts to improve early childhood financing. They have undertaken a series of related activities to build a responsive support and technical assistance capacity to help "reform-ready" states and communities to advance their agendas:

- Clarify issues that are unique to financing community supports and services for young children and their families;
- Identify and assess promising reform strategies;
- Develop and distribute relevant policy tools.

III. Promotion of improvements in financing systems and strategies.

The Finance Project reaches out to an array of audiences that are critical players in the change process—including federal, state, and local officials; educators, human service providers, and community organizers in non-profit and quasi-governmental organizations; business and union leaders; academic scholars and foundation executives; and the media. The Finance Project has created a valuable central intelligence capability on financing issues and strategies, as well as a rich support and technical assistance resource for "reform ready" states and communities. Drawing on the knowledge and tools that are being produced, the project's staff, consultants, and working group members work directly with state and community leaders in several ways:

- Facilitating direct connections with peers and experts who have relevant knowledge, expertise, and experience.
- Providing on-site learning opportunities.
- Producing tailored "how to" materials.
- Creating learning clusters.'

In September 1997, The Finance Project was awarded a contract from the Child Care Bureau of the U.S. Department of Health and Human Services to work with the Families and Work Institute and the National Governor's Association to provide information and technical assistance to state child care administrators as they work with businesses, philanthropic organizations, and other groups to build and sustain partnerships. Initial activities will

- Identify and track existing practices for promoting public-private partnerships at both the state and community levels;
- Develop a series of written technical assistance materials to guide the formation and implementation of successful public-private partnerships; and
- Provide periodic updates to state child care grantees about information collected and produced.

The Finance Project extends the work of many other organizations and blue-ribbon groups that have presented bold agendas for improving supports and services for children and families. It is creating the vision for a more rational approach to generating and investing resources in education, other supports and services for children and families, and communities. It is developing ideas, options, and policy tools to actively foster positive change through broad-based systemic reform, as well as through more incremental steps to improve the effectiveness, efficiency, and equity of current systems. It also provides support and technical assistance to "reform ready" states, communities, and initiatives engaged in efforts to align their financing systems with their policy and program reform agendas.

For more information about The Finance Project and its activities, please contact:

Cheryl D. Hayes, Executive Director
The Finance Project
1000 Vermont Avenue, NW, Suite 600
Washington, DC 20005
202/628-4200
202/628-4205 (Fax)
info@financeproject.org (E-mail)

RESOURCES AVAILABLE FROM THE FINANCE PROJECT'S WORKING PAPERS SERIES

Federal Financing Issues and Options

_____ *Financing Services for Young Children and Their Families: Meeting the Challenges of Welfare Reform* by Cheryl D. Hayes (March 1997)

_____ *Federal Tax Reform: A Family Perspective* by Michael J. McIntyre and C. Eugene Steuerle [Report and Executive Summary] (July 1996) *Please note there is a $10.00 charge for this publication.

_____ *The Budget Enforcement Act: Implications for Children and Families* by Karen Baehler (November 1995)

_____ *Dollars and Sense: Diverse Perspectives on Block Grants and the Personal Responsibility Act* (Joint publication of The Finance Project and the American Youth Policy Forum and The Policy Exchange of the Institute for Educational Leadership) (September 1995)

_____ *Rethinking Block Grants: Toward Improved Intergovernmental Financing for Education and Other Children's Services* by Cheryl D. Hayes, with assistance from Anna E. Danegger (April 1995)

_____ *Reform Options for the Intergovernmental Funding System: Decategorization Policy Issues* by Sid Gardner (December 1994)

State Financing Issues and Options

_____ *Money Matters: A Guide to Financing Quality Education and Other Children's Services* (January 1997) *Please note there is a $20.00 charge for this publication.

_____ *The Effects of Economic and Demographic Changes on State and Local Budgets* by Sally Wallace (December 1995)

_____ *Issues and Challenges in State and Local Finance* by Therese J. McGuire (November 1995)

_____ *Toward State Tax Reform: Lessons From State Tax Studies* by Therese J. McGuire and Jessica E. Rio (November 1995)

_____ *Legal Issues and Constraints Affecting Finance Reform for Education and Related Services* by Thomas Triplett (November 1995)

_____ *State Investments in Education and Other Children's Services: The Fiscal Challenges Ahead* by Martin E. Orland and Carol E. Cohen (November 1995)

_____ *State Investments in Education and Other Children's Services: Fiscal Profiles of the 50 States* by Steven D. Gold, Deborah A. Ellwood, Elizabeth I. Davis, David S. Liebschutz, Sarah Ritchie, Martin E. Orland, and Carol E. Cohen (October 1995)

_____ *State Investments in Education and Other Children's Services: Case Studies of Financing Innovations* by Ira M. Cutler, Alexandra Tan, and Laura Downs (September 1995)

_____ *Spending and Revenue for Children's Programs* by Steven D. Gold and Deborah A. Ellwood (December 1994)

Local Financing Issues and Options

_____ *Tax Strategies for Community Economic Development* by Paul Pryde, Jr. (June 1998)

_____ *Money Matters: A Guide to Financing Quality Education and Other Children's Services* (January 1997) *Please note there is a $20.00 charge for this publication.

_____ *The Property Tax in the 21st Century* by Hal Hovey (May 1996)

_____ *Issues and Challenges in State and Local Finance* by Therese J. McGuire (November 1995)

Financing Comprehensive, Community-based Supports and Services and Improving Service Delivery

_____ *Privatization, Contracting, and Reform of Child and Family Social Services* by Sheila B. Kamerman and Alfred J. Kahn (June 1998)

_____ *Developing Cost Accounting and Decision Support Software for Comprehensive Community-Based Support Systems: An Analysis of Needs, Interest, and Readiness in the Field* by Robert Harrington and Peter Jenkins with Carolyn Marzke and Carol Cohen (June 1998)

_____ *Financing Strategies to Support Comprehensive, Community-based Services for Children and Families* by Mary M. O'Brien, National Child Welfare Resource Center for Organizational Improvement (March 1997)

_____ *Building Strong Communities: Crafting a Legislative Foundation* (December 1996) *Please note there is a $20.00 charge for this publication.

_____ *Building Comprehensive, Community-based Support Systems for Children & Families: A Review of Legislative Examples* by Thomas Woods (December 1996)

_____ *Beyond Decategorization: Defining Barriers and Potential Solutions to Creating Effective Comprehensive, Community-based Support Systems for Children and Families* by Martin E. Orland and Ellen Foley (April 1996)

_____ *Conceptualizing the Costs of Comprehensive, Community-based Support Systems for Children* by Jennifer King Rice (November 1995)

_____ *Creating More Comprehensive, Community-based Support Systems: The Critical Role of Finance* by Martin E. Orland, Anna E. Danegger and Ellen Foley (November 1995)

_____ *Compendium of Comprehensive, Community-based Initiatives: A Look at Costs, Benefits, and Financing Strategies* by Cheryl D. Hayes, Elise Lipoff, and Anna E. Danegger (July 1995)

_____ *The Role of Finance Reform in Comprehensive Service Initiatives* by Ira M. Cutler (December 1994)

Results-based Planning, Budgeting, Management and Accountability Issues

_____ *A Guide to Developing and Using Performance Measures* by Mark Friedman (May 1997)

_____ *A Guide to Results and Indicators* by Atelia Melaville (May 1997)

_____ *A Strategy Map for Results-based Budgeting: Moving from Theory to Practice* by Mark Friedman (September 1996)

Forthcoming

A Guide to Developing Child and Family Budgets by Mark Friedman and Anna E. Danegger (July 1998)

Results-based Planning, Budgeting, Management, and Accountability Strategies: An Annotated Bibliography by Anna E. Danegger (Summer 1998)

Financing Early Childhood Supports and Services

_____ *Financing Services for Young Children and Their Families: New Directions for Research, Development, and Demonstration* (June 1998)

_____ *Revenue Generation in the Wake of Welfare Reform: Summary of The Pilot Learning Cluster on Early Childhood Finance* (August 1997)

_____ *Financing Services for Young Children and Their Families: Meeting the Challenges of Welfare Reform* by Cheryl D. Hayes (March 1997)

School Finance Issues

____ *Securing Equal Educational Opportunities: Past Efforts and the Challenges Ahead* by Alexandra Tan and Martin E. Orland (February 1995)

____ *School Finance Litigation: A Review of Key Cases* by Dore Van Slyke, Alexandra Tan and Martin E. Orland, with assistance from Anna E. Danegger (December 1994)

Working papers produced by The Finance Project cost $7.50. However, *Federal Tax Reform: A Family Perspective* [Report and Executive Summary] is $10.00. In addition, *Money Matters: A Guide to Financing Quality Education and Other Children's Services* and *Building Strong Communities: Crafting a Legislative Foundation* cost $20.00 each. Please indicate the quantity desired next to the publications you would like to receive and mail this form, along with your payment, to:

The Finance Project
1000 Vermont Avenue, NW, Suite 600
Washington, DC 20005
202/628-4200

Name: _____

Title: _____

Affiliation: _____

Address: _____

_____

City: _____ State: _____ Postal Code: _____

Telephone: _____ Fax: _____

Total Number of Publications Ordered: _____ Total Cost: _____

PLEASE MAKE CHECK PAYABLE TO THE FINANCE PROJECT. PREPAID ORDERS ONLY.
Federal Tax ID# 52-184-1608 July 1998